AUGUST
1 9 5 3
VOLUME 17
NUMBER 4

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Edited by Laurance F. Shaffer

Teachers College, Columbia University

Associate Editors: Edgar A. Doll, *Devereux Schools, Devon, Pa.* William A. Hunt, *Northwestern University.* E. Lowell Kelly, *University of Michigan.* Morris Krugman, *Board of Education, New York, N. Y.* Bertha M. Luckey, *Board of Education, Cleveland, Ohio.* Ann Magaret Garner, *University of Illinois.* Fred McKinney, *University of Missouri.* Anne Roe, *La Jara, N.M.* Carl R. Rogers, *University of Chicago.* R. Nevitt Sanford, *University of California.* Edward Joseph Shoben, Jr., *Teachers College, Columbia University.* J. R. Wittenborn, *Yale University.*

Lorraine Bouthilet, *Managing Editor*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy, and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problem of clinical psychology also appear in its pages.

Editorial Office

Manuscripts submitted for publication, books and tests for review, and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y.

Information for contributors: The important characteristics of a good article are the significance of its subject matter, the demonstrated soundness of its conclusions, its clarity, and its brevity. Clarity is improved by careful outlining, discriminating use of headings, and a simple, straightforward style of writing. For brevity, authors should strive to convey every thought without waste of words, and should avoid repetitions of methods, data or findings.

The style of all manuscripts should conform to the instructions given in the Publication Manual of the American Psychological Association, prepared by the Council of Editors (*Psychological Bulletin*, 1952, 49, 389-449). If it is not otherwise available, the *Manual* may be purchased, for $1.00, from the Washington office of the Association. The Editor expects each author to study the *Manual* carefully before submitting a manuscript. Special attention should be given to the style of headings, capitalization, indention, spacing, and footnotes, and to the preparation of tables, figures, and lists of references. Otherwise acceptable manuscripts which depart seriously from the standards of the *Manual* are returned to the author for revision and retyping.

There is a lag of about nine to twelve months between the acceptance of an article and its publication. Authors may secure immediate publication, however, by arranging for articles to be printed as extra pages, increasing the number of pages that subscribers receive. The cost to the author of such early publication is approximately $14.00 per printed page.

Authors receive fifty reprints gratis, and may order additional quantities when returning proof. Authors are charged one-half of the extra cost of tabular material and cuts, and the full cost of changes made on the proof.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St., N. W., Washington 6, D. C.

The Journal of Consulting Psychology is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $7.00, foreign $7.50, single number $1.25.

Copyright, 1953 by the American Psychological Association, Inc. Printed and mailed by the Dentan Printing Co., 23 West Colorado Ave., Colorado Springs, Colo. and entered as second-class matter at the Post Office at Colorado Springs, Colo. Acceptance for mailing at special rate of postage provided for in section 538, Act of February 25, 1925, authorized May 27, 1940.

JOURNAL OF CONSULTING PSYCHOLOGY

August, 1953 Vol. 17, No. 4

CONTENTS

Some Test Correlates of Response to Psychotherapy: *Frank Barron* - - - - - - - 235

A Nonintellectual Intelligence Test: *Harrison G. Gough* - - - - - - - - - - 242

A Repetition of Meyer Williams' Study of Intellectual Control under Stress and Associated Rorschach Factors: *Virgil R. Carlson and Richard S. Lazarus* - - - 247

Subcategories of Rorschach Human Movement Responses: A Classification System and Some Experimental Results: *Robert P. Barrell* - - - - - - - - - - - - - 254

Further Investigation of Rorschach Determinants Subjected to Factor Analysis: *Harold L. Williams and James F. Lawrence* - - - - - - - - - - - - - 261

The Effect of Test Anxiety on Memory Span in a Group Test Situation: *Donald Kaye, Paul Kirschner, and George Mandler* - - - - - - - - - - - 265

Neurocirculatory Reactions and a Group Stress Situation: *George B. Strother and Donald M. Cook* - - - - - - - - - - - - - - - 267

Responses of Women to the Thematic Apperception Test: *Leonard D. Eron* - - - - 269

Ambiguity of Pictures and Extent of Personality Factors in Fantasy Responses: *Douglas T. Kenny and Sidney W. Bijou* - - - - - - - - - - - - - - - 283

Measuring Achievement Motivation with "Picture Interpretations": *Henry H. Morgan* 289

Social and General Intelligence and Performance on the Rosenzweig Picture-Frustration Study: *Lawrence Karlin and Milton M. Schwartz* - - - - - - - 293

The Practical Utility of the Clinical Report: *Charles A. Dailey* - - - - - - - - 297

Clinical Judgment and the Cattell Infant Intelligence Scale: *James J. Gallagher* - - 303

Three Factors in Parent Behavior: *Maurice Lorr and Richard L. Jenkins* - - - - 306

Follow-up Study of Ten-Year-Old Boys with Police Records: *William W. Wattenberg and Frank Quiroz* - - - - - - - - - - - - - - - 309

A Note on the *K* Scale of the Minnesota Multiphasic Personality Inventory: *Anders Sweetland and Herbert Quay* - - - - - - - - - - - - - - - - - 314

The Archimedes Negative Aftereffect as an Indication of Memory Impairment: *Lloyd S. Standlee* - 317

New Books and Tests - 318

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Some Test Correlates of Response to Psychotherapy[1]

Frank Barron
Institute of Personality Assessment and Research
University of California

This paper reports certain personality and intelligence test correlates of changes in psychological morbidity in 33 adult psychoneurotics following six months of psychotherapy in the outpatient service of a state psychiatric clinic. These findings are part of the results of a study of somewhat broader scope, in which psychotherapy was considered as a special case of personal interaction in general, and in which therapeutic relationships and outcomes were rationalized in terms of the personality structures of therapists as well as of patients [1].

The present paper, in addition to being restricted to a report on the test results only of patients, will adopt by choice a somewhat narrow conception of response to psychotherapy. The criterion variable can be best characterized as an answer to the question, "Did the patient improve or not?" where the word "improve" means some fairly general changes in state from bad to good. Examples of such changes are these: (*a*) The patient feels better, is more comfortable, takes more interest in life, and the like; (*b*) important interpersonal relations are straightened out a bit; (*c*) physical symptoms have been relieved or cured; (*d*) important health-tending decisions have been made; (*e*) there has been an increase in insightful remarks and behavior.

The judgment as to whether or not improvement had occurred was made by two expert observers (heads of departments at Langley Porter Clinic) who had not had any part in the conduct of therapy. Each therapist made a formal presentation before these two judges of every case he had handled; prior to the presentation the judges had read all of the material concerning the patient which had been recorded in the clinical chart.

In the instruction to the judges it was emphasized that the crucial variable was *not* general level of functioning of the patient at the conclusion of the therapy, but rather the *change in state* which had occurred between beginning and end of therapy. Further, it was made clear to the judges that part of their function was to evaluate the therapist's involvement in his own account of the therapeutic process, and to weigh that factor in coming to a best estimate as to the degree of change which had actually occurred.

On the basis, then, of two main sources of information (formal presentation of the case by the therapist, and an evaluation from the clinical chart), the expert raters assigned cases, first of all, into two main categories, those who had shown definite improvement and those who had failed to improve or who had improved only slightly. They were in close agreement on this crude categorization, differing on only 2 of the 33 cases which were studied and rated. In addition, they assigned ratings on a 100-point scale to each case, so that the total sample was ranked in terms of the variable "Improvement." The correlation between the two independent sets of ratings was .91. The two disagreements were resolved by averaging the scale ratings. The final classification included 17 patients in the Improved group, 16 in the Unimproved.

Some Illustrations of Criterion Behavior

It is of some interest to examine the kinds of incidents and outcomes which the judges con-

[1]This study was conducted for the most part at Langley Porter Clinic, San Francisco. It was completed at the Institute of Personality Assessment and Research, and was supported financially by the grant of the Rockefeller Foundation to that Institute. The author is indebted to Robert E. Harris, director of the psychology service at Langley Porter, for many valuable suggestions which contributed to the research.

sidered indicative of improvement or lack of it. Here is a partial list of such incidents and outcomes, abstracted from the clinical charts of these patients.

Improvement:

1. A woman patient who had been frigid through four years of marriage which terminated in divorce now established an intimate relationship with another man in which for the first time she experienced orgasm in sexual intercourse.

2. A man who entered therapy in a very depressed, anxious, and physically upset state, and whose troubles centered on his relations with his foreman on a construction job (a relationship in which he was outwardly submissive and cooperative but inwardly enraged), finally learned to "stand up to" the foreman and express his feelings. There was a clear advance in his feelings of independence and self-esteem, and towards the conclusion of the therapy the patient left his former job and started a business of his own.

3. A woman with menstrual difficulties related to unconscious hostility towards her husband because of his disregard of her own wishes in sexual intercourse became conscious both of her anger and her retaliative tactics, with consequent disappearance of the menstrual difficulties and a more candid and unmartyred relationship to her husband.

4. A woman who had been recently divorced after an extremely traumatizing marriage and who entered therapy in a state of tremendous agitation and anxiety, so incapacitating that she could not sit in the lobby of the clinic or cross a street or look anyone in the face, gradually lost her phobias, experienced a great reduction in anxiety, and became able to associate comfortably with other people.

5. A man suffering from severe gastric disturbances which he considered unrelated to any emotional problems entered therapy upon the strongly worded advice of the internist. His dissociation of affect from the death of his mother and subsequent disturbing childhood years in an orphanage was gradually broken down and he reexperienced his own sense of forsakenness and finally came to terms with it, crucially aided by the support and understanding of the woman therapist.

Lack of improvement:

1. A very dogmatic and dominating man who entered therapy at the insistence of his wife (who complained of his overbearingness with her and their three children) proved too overbearing for the young woman therapist as well. The therapist was unable to handle the negative countertransference which developed, and the therapeutic interaction became simply a battle, with consequent lack of improvement in the patient.

2. A woman of forty, who had never been married and who suffered from continual headaches and disturbing sexual impulses which she attempted to suppress, found herself obliged to terminate therapy for some good reason precisely after the hour at which the therapist had told her he would give her a physical examination on the occasion of their next appointment.

3. A man with a history of homosexuality attempted to seduce his male therapist, who responded with anger as well as some anxiety. The patient did not return.

4. A psychopathic young man who had a long history of rebelliousness, brushes with the law, and employment instability entered therapy because, he said, he happened to be passing by the clinic and wondered what it would be like to be a patient. He liked it for a while, especially as he happened to be assigned to an attractive woman therapist, but after a few months he hitchhiked off in a southerly direction, leaving a chatty farewell note for the doctor in which he addressed her by her first name.

5. An engineer of extremely masculine appearance and manners complained of a "psychic malfunction" and wanted "a few treatments to get at the nuclear material and get the thing cleared up." His therapist insisted on taking a nondianetic, unhurried approach to the matter, which annoyed the patient, who nevertheless remained in therapy in the expectation that something would eventually happen.

One might summarize the kinds of outcomes in the Improved group as follows: (*a*) The patient generally reported a feeling of well-being at the conclusion of therapy, in contrast to depression and anxiety at its start; (*b*) specific symptoms, such as headaches, frigidity, or impotence, gastric disturbances, menstrual difficulties, skin disorders, etc., tended to be relieved or totally cleared up; (*c*) in some cases there were significant changes in the direction of more mature interpersonal relations, especially with parents, parent-substitutes, or spouses.

The failures in psychotherapy could often be traced to the inability of the therapist to handle some particularly difficult problem. Perhaps with more experienced psychotherapists some of the patients who did not improve would have made some progress; however, there seemed little doubt that the cases which were marked down as therapeutic failures were basically more difficult problems. As we shall see later from the test results, it was generally the more disturbed individuals who did not improve.

Test Correlates of Improvement

The test battery given these patients before

psychotherapy began consisted of a shortened form of the Wechsler-Bellevue Intelligence Scale, the Minnesota Multiphasic Personality Inventory, The Rorschach Psycho-diagnostic and the Ethnocentrism Scale (in Form 60 of the University of California Public Opinion Study scales). The patients themselves were all receiving psychotherapy for the first time, and they began the course of treatment just at the time that their therapists were starting a six-month period of service in the outpatient clinic. The therapists were psychiatric residents who had had little experience in intensive psychotherapy.

Most of the patients had one hour of treatment a week for six months. Of the 33 patients, 12 were men and 21 were women. They ranged in age from 20 to 45. All but 3 of them were high school graduates, but only 2 were college graduates. As a group they could probably be characterized as lower-middle-class socioeconomically. Two-thirds of them were gainfully employed.

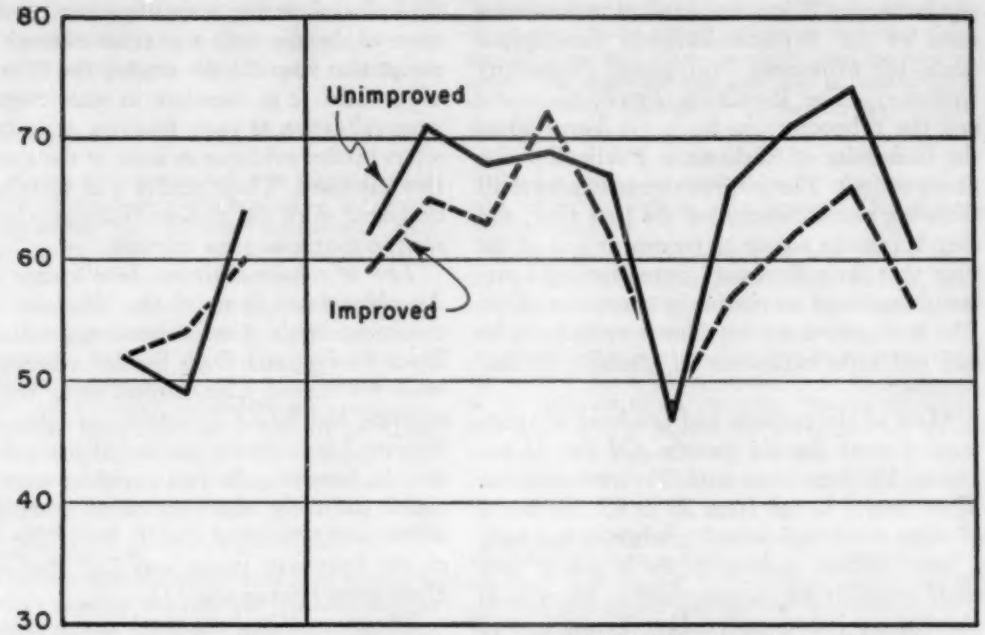
All of the testing was done by the experimenter. Conventional controls were exercised to prevent contamination in the prediction of outcome from the tests. The controls were as follows: (*a*) Only the experimenter knew the test results. The psychotherapists were informed of this condition before the experiment began, and were told that they would be given information concerning the testing only if in their judgment it was essential to the diagnostic formulation and the handling of the case. In three cases such information was requested by the therapist during the course of the therapy. The information given was not comprehensive, but was in response to some specific point. These cases were retained in the sample in spite of their being exceptions to the general rule, for their exclusion would be as likely a source of contamination as their retention. (*b*) The judges of outcome had no knowledge of the test results. (*c*) The test protocols were identified by a code number rather than by name, so that clinicians who attempted to predict outcome had no knowledge of the patient beyond that afforded by the test results themselves.

The design of the present study, so far as prediction of outcome is concerned, is a replication of some earlier work by Harris and Christiansen [5], who used in an almost identical clinical setting a similar criterion of outcome of therapy and the same clinical tests, except that they did not employ the Ethnocentrism scale. It is therefore in some respects a cross-validation of their findings, or at least it offers further evidence on some of the questions they discussed. Their results will therefore be compared with the present findings whenever such comparisons seem relevant.

The Wechsler-Bellevue Intelligence Test. An abbreviated form of the Wechsler scale, consisting of the Comprehension, Similarities, Block Design, and Digit Symbol subtests, was used. Harris and Christiansen, using the same subtests, had found no significant relationship between Improvement and Intelligence. In this sample, however, the two variables were associated positively, the Pearsonian r being .46, which is significant at the .01 level. The mean of the Improved group was 122, that of the Unimproved group was 112.

The mean IQ of the total sample was 117, as compared with a mean of 108 in the Harris-Christiansen study. An important difference between the two samples so far as method of selection is concerned should be noted: The Harris-Christiansen sample consisted of already hospitalized persons who were given psychotherapy because of delayed recovery from physical disease, surgery, or accident, whereas almost all of the patients in the present study had elected on their own initiative to seek psychotherapy at the clinic, and in their waking hours went about the business of life in an upright position. This difference might well account for the IQ difference between the samples, as well as for the fact that these self-referred patients were a good standard deviation above the general population mean in intelligence. Simply being aware of the fact that psychotherapy is to be had and that it makes sense to seek it when you are in personal difficulties is probably related positively to general intelligence and cultural sophistication.

It is also to be expected on theoretical grounds that greater effectiveness of cortical functioning (which presumably reveals itself on the Wechsler-Bellevue) should be associated with a factor of modifiability in personality structure. Intelligence certainly involves the ability to cognize relationships adequately, including emotional relationships, and to correct



	L	K	F	Hs	D	Hy	Pd	Mf_m	Mf_f	Pa	Pt	Sc	Ma
Improved	52	54	60	59	65	63	72	62	48	57	62	66	56
Unimproved	52	49	64	62	71	68	69	67	47	66	71	74	61

Fig. 1. Group MMPI profiles of the Improved and Unimproved groups.

one's cognitions on the basis of new evidence. It would seem that in this sample, at any rate, the more intelligent patients were better able to use the psychotherapeutic relationship to induce desired personality changes.

The Minnesota Multiphasic Personality Inventory. The MMPI analysis may be divided into three main sections: (*a*) group differences between the Improved and Unimproved cases on MMPI scales; (*b*) prediction of outcome from MMPI profiles by clinical psychologists skilled in the use of the test, as well as by certain rational indices or rules for mechanical sorting of the profiles; (*c*) development and cross-validation of a prediction scale. Enterprises 1 and 2 will be reported in this paper; the development and cross-validation of a prediction scale will be reported separately, however, as it involved work with a number of other samples and raised some test-specific questions which are not properly within the scope of this paper [2].

Group comparisons on the usual diagnostic and validity scales are shown in Figure 1. Only on the Paranoia scale was there a significant difference between the means of the Improved and Unimproved groups (t of 2.6, $p < .02$). The Unimproved group profile is consistently higher on almost all scales, however, and the peak is at Schizophrenia, with the average T scores for Psychasthenia and Depression also being greater than 70. The sorts of individuals who earn profiles like this are usually schizoid; in *An Atlas for the Clinical Use of the MMPI* [6], most of the 8724 profiles carry a diagnosis of reactive depression with underlying schizoid trends (pp. 668-669). This fits well with the observation by Harris and Christiansen that poor prognosis is generally associated with what they call "subclinical psychotic trends," referring to a more severe ego dysfunction than first clinical impression seems to indicate. Depression and anxiety are salient features of the clinical picture in almost all patients who seek psychotherapy in the outpatient clinic; but the patients who fail to improve are those in whom paranoid and schizoid features underlie the psychoneurotic symptoms. A reasonable guess

would be that their personal difficulties are more chronic and characterologically based, in contrast to the more acute and situation-linked problems of the patients who improve.

The 33 individual MMPI profiles were also given to clinical psychologists for prognostic evaluations. Eight clinicians independently attempted to predict outcome from the profiles, having no other information concerning the patient beyond age and sex. Each clinician sorted the 33 records into two groups, predicting either that the case would be improved or that it would be unimproved. Of 264 (8×33) such classifications, 164 (62.12%) were correct. The sum of the eight values of chi square was 20.46, which for 8 *df* yields a p less than .01. The three best sorters had a pooled accuracy percentage of 69.7%. All of the sorters were more often correct than in error, but it should be pointed out that an average accuracy of 62% when 50% would be expected by chance is a fairly modest achievement in clinical prediction. Even the best of the MMPI analysts would be well advised to be tentative in their prognostic formulations.

Rational indices and rules for mechanical sorting produce substantially the same accuracy of prediction as do sortings by clinicians; actually, such indices are slightly more effective. Two fairly simple and obvious mechanical predictors were put to the test. The first of these employed the rule that for any profile with all scores within the normal range (i.e., greater than 30, less than 70) Improvement should be predicted. Ten profiles met this requirement, and 9 of these cases had in fact improved. Twenty-four of 33 profiles (73%) were thus classified correctly (chi square of 8.1, $p<.01$). Exceptions were then allowed in the general rule: e.g., profiles with only D, or with only D and *Pt*, greater than 70 would be called Improved. Both exceptions led to an accuracy of prediction between 75 and 80%.

In addition to these sorting rules, a simple index of subclinical psychotic trends was defined by summing the scores of four scales (*F, Pa, Sc,* and *Ma*) usually found elevated in psychosis. This index correlated .40 with outcome as it had been rated on a 100-point scale.

It is of some interest that the same indices which predicted response to psychotherapy worked equally well in another setting, with a sample apparently quite different from the patient sample at this clinic. In a study of personal effectiveness in male graduate students in their final year of work for the doctorate, ratings were obtained from faculty members on certain dimensions such as originality and personal soundness. The averaged faculty ratings on personal soundness for 40 such students correlated—.4 with the index defined by the sum of *F, Pa, Sc,* and *Ma,* and had a contingency of .52 with a mechanical sorting of the MMPI profiles using the rule that all profiles having no clinical scale elevation (except *D*) of 70 or above should be called Sound, while all others should be called Unsound.

The "personally sound" graduate students were different from the "personally unsound" students in the same way that patients who profited considerably from psychotherapy differed from patients who profited little or not at all. As one would expect, however, there was a difference in absolute level of adjustment; the distribution of scores on the Prognosis for Psychotherapy scale [2] showed a clear progression in group means from the Unimproved patients through the Improved patients and the Unsound graduate students to the Sound graduate students.

What these relationships indicate, it would seem, is that the patients who are most likely to get well are those who are not very sick in the first place. Another way of putting this is to say that patients who are more integrated to begin with are better able to use the psychotherapeutic relationship to solve whatever problems brought them into therapy. A corollary of this is that the potential gains from therapy should increase as therapy progresses, at least up to the point where the critical problem for the patient is to become genuinely independent of the therapist. By that time, of course, the therapeutic process has become internalized, and personality problems have been brought "into the ego," to be dealt with rationally and objectively.

The Rorschach Psychodiagnostic. The Rorschach analysis was carried out according to the same general design as the MMPI analysis, with the exception that no prognostic scale was developed. Average profiles were determined for the Improved and Unimproved groups, using the Klopfer-Kelly conventions for the scor-

ing and nomenclature of determinants. Individual scores on the Harris-Christiansen Rorschach Prognostic Index [5] were computed, and the relationship of Outcome to that index was calculated. Finally, the test protocols, including both content of responses and profiles of determinants, were given to clinicians for sorting.

All of Rorschachs were administered by the experimenter, with all of the verbalizations and relevant behavior of the patient being recorded. The tests were scored independently by the experimenter and by one other clinical psychologist who had had several years of experience both in the scoring and the interpretation of the Rorschach. The experimenter himself had had approximately three years of clinical experience with the test, in addition to a graduate course in its scoring and interpretation. The two independent scorings showed a high degree of agreement; discrepancies were reconciled in a conference on review of the scoring. The Improved and Unimproved groups did not differ significantly on any Rorschach determinant, nor on any of the important ratios, such as *W*, *D*, *Dd*, plus *S*, *M*: sum *C*, and so on. This finding is consistent with Rorschach results in the gradute student sample previously referred to; in that study no significant relationship was found between any Rorschach determinant or combination of determinants and the criterion variable, Personal Soundness.

The Harris-Christiansen Prognostic Index correlated .00 with Improvement. As Cronbach [3] has pointed out, most empirical Rorschach indices fail to hold up under cross-validation, so that this is not a surprising finding.

Somewhat contrary to the expectation of most persons who have worked with the test, however, is the finding that experienced Rorschach interpreters were not able to predict outcome of therapy on the basis of the Rorschach evaluation. Four clinicians, all of whom were highly trained and experienced, both in the use of the test and in general clinical practice, attempted the sorting task. They had at hand all of the responses of the patients, as well as a summary of the scoring in terms of the usual ratios and profile of determinants. Only one of the four sortings approached significance, and that was in the negative direction; that is, the sorter was more often wrong than right in predicting outcome.

The Ethnocentrism scale. Certain theoretical considerations had led to the inclusion of the Ethnocentrism scale in the test battery. Levinson [7] had found, in a study at this same clinic, that ethnocentric patients showed a personality syndrome of a sort which would be resistant to change through psychotherapy. High-scoring patients were described by her as having the following characteristics: "rigid, constricted personalities . . . stereotyped, conventionalized thinking . . . undifferentiated egos . . . narrow range of experience, emotionally and intellectually . . . weak interpersonal relationships, expressed in terms of dominance-submission." The low-scoring patients, on the other hand, were marked by their "desire to be loved and fear of being rejected . . . expressed directly in interpersonal relations and in a very personalized way . . . the frequency with which low scorers discuss their relationships to others is striking."

These considerations suggested the hypothesis that patients who scored high on the Ethnocentrism scale would be less likely to enter deeply into the sort of personal interaction which occurs in psychotherapy. From the very beginning of therapy they would find it necessary to narrate the comings and goings of their somatic complaints, and to minimize the importance of their psychological problems as the cause of their somatic disturbances. According to this notion, ethnocentric patients would tend characteristically to isolate affect and to avoid psychological self-examination; they would be strangers to their own inner life.

While the psychological formulation itself was not checked on in any detail in the present study, the hypothesis that it led to (i.e., that Ethnocentrism would be negatively related to Improvement) was confirmed. The correlation of the *E* scale with Improvement ratings was —.64, making it the best single predictor employed in this study.

Several reservations, however, should be noted. In the first place, scores on the *E* scale are negatively related to intelligence, which can be presumed to be a somewhat more basic psychological variable than ethnocentrism. As Gough [4] has pointed out, many of the items in the *E* scale are written in such an extreme fashion that a person of some intelligence and

education could not agree with them, even though the sentiment expressed might win his approval if the form of its statement were less clearly irrational. Since Improvement is correlated positively with Intelligence in this sample (.46), one would expect a negative relationship between Improvement and Ethnocentrism on the basis simply of their common relationship to Intelligence.

When Intelligence is partialled out, however, some relationship remains. The coefficient of partial correlation between Ethnocentrism and Improvement, with Intelligence held constant, is—.34. Partialling intelligence out in this manner is not strictly correct, of course, except in a mechanical way; actually, some underlying emotional-cognitive variable such as rigidity, which is complexly related both to intelligence and to ethnocentrism, may be accounting for a large part of the correlation of both of those measures with the capacity for psychological change as shown by improvement in therapy.

There is one other reservation to be noted. Four of the psychotherapists in this study were Jewish, a fact which might have affected both the ethnocentric patient's attitude towards them and their attitude towards the ethnocentric patient. In several of the therapeutic interactions, such patient attitudes came out into the open, and were dealt with therapeutically; and in at least one case the therapist admitted candidly to the psychologist that his own response to the patient's anti-Semitism was disruptive of the relationship.

In spite of these reservations, however, the relationship between ethnocentrism and capacity for psychological change through psychotherapy is a striking one, and if confirmed would be of some theoretical importance. It may, indeed, have implications for social communities as well as individuals, and would suggest that the social organization which fosters prejudice might well be forfeiting the kind of emotional flexibility which is necessary if in time of crisis it is to cure its own ills.

Summary

Change in psychological morbidity as a consequence of six months of psychotherapy was studied in a sample of 33 adult psychoneurotics who were treated in the outpatient service of a state psychiatric clinic. Changes for the better might be summarized as follows: (*a*) The patient generally reported a feeling of well-being at the conclusion of therapy, in contrast to depression and anxiety at its start; (*b*) various psychosomatic symptoms tended to be relieved or cured; (*c*) significant interpersonal relations were established on a more mature basis.

Patients were rated on the degree of their improvement by two skilled judges who evaluated both the interview-by-interview account of the therapy in the clinical chart and the therapist's report in conference on the course and outcome of the therapy. Ratings of degree of change for the better were then correlated with various psychological tests.

The chief test correlates are these: (*a*) Intelligence is positively associated with change; (*b*) level of integration at the beginning of therapy is positively associated with change (i.e., those who are better off to begin with are likely to improve the most); (*c*) ethnocentrism is negatively related to change.

Received December 11, 1952.

References

1. Barron, F. Psychotherapy as a special case of personal interaction: prediction of its course. Unpublished doctor's dissertation, Univer. of California, 1950.
2. Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, in press.
3. Cronbach, L. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.*, 1949, **46**, 393–429.
4. Gough, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 237-246.
5. Harris, R. E., & Christiansen, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, **21**, 269-284.
6. Hathaway, S. R., & Meehl, P. E. *An atlas for the clinical use of the MMPI.* Minneapolis, Minn.: Univer. of Minnesota Press, 1951.
7. Levinson, Maria H. Psychological ill-health in relation to potential fascism: a study of psychiatric clinic patients. In T. W. Adorno, *et al.*, (Eds.) *The authoritarian personality.* New York: Harper, 1950. Pp. 891-970.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

A Nonintellectual Intelligence Test [1, 2]

Harrison G. Gough
University of California, Berkeley

The title of this paper may be something of a misnomer, for it is not at all certain that the testing instrument which is going to be presented is, actually, a measure of intellect. Nevertheless, it does correlate significantly with conventional tests of intelligence, and seems to reveal other properties which suggest that something along the line of intellectual ability, or perhaps more precisely intellectual efficiency, is being appraised. The other part of the title—the "nonintellectual" part—has not been the cause of any misgivings, because, as will be seen shortly, the test instrument does not call for any problem solving, reasoning, abstract thinking, or any of the other mental functions which are traditionally regarded as the key factors in defining and measuring intelligence.

The two immediate aims of this project were first, to determine whether a criterion-specific set of personality inventory items could be assembled which would correlate significantly with accepted measures of intellect, and second, if this was successful, to determine whether the individual items could be combined in a scale which would have anything approaching practical significance and value.

The second objective is added, because, as we know for instance from work on the Rorschach [1], a series of variables can easily be identified which will correlate significantly with intellect, but which will seldom rise above the level of + .20 and + .25 into a region of more dependable utility when pooled in a composite.

Procedures

The first step in the present project was to write a series of criterion-specific inventory items, and also to select appropriate items from standard sources. An example of a spe-

[1] This paper is one of a series of studies concerned with the development of practical assessment devices for the identification of positive and favorable aspects of personality and temperament. The project is being carried out under a research grant from The National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service.

[2] The present paper is a revision and extension of a preliminary report given at Fresno, California, April 26, 1952 at the annual meeting of the Western Psychological Association.

Table 1
Item Statistics for Four Items from the *Ie* Scale

Item	Percent saying "True"					
	Higher scores on intellectual ability measures			Lower scores on intellectual ability measures		
	School A (*N* 28)	Schools B and C (*N* 21)	School D (*N* 48)	School A (*N* 28)	Schools B and C (*N* 21)	School D (*N* 48)
1. It is more important that a father be kind than that he be successful.	79	95	92	71	76	69
2. Parents are much too easy on their children nowadays.	36	43	46	50	71	67
3. I am not afraid of picking up a disease or germs from door knobs.	86	81	85	79	67	67
4. Most people make friends because friends are likely to be useful to them.	21	19	29	46	71	77

Table 2
Summary Statistics for the First (76-item) Intellectual Efficiency (*Ie*) Scale in the Original Samples

Sample	*N*	*M*	*SD*	*r**
School A	135	52.44	8.51	.47
School B	35	51.34	9.20	.68
School C	66	51.41	9.41	.64
School D	205	51.93	9.51	.54

* Intercorrelations between the Ie scale and a conventional IQ test in each of the schools.

Table 3
Summary Statistics for the First *Ie* Scale in Six Cross-Validating Samples *

Sample	*N*	*M*	*SD*	*r*†
I. For the abbreviated (59-item) scale.				
1. St. Cloud ninth grade	82	38.09	7.46	.53
2. Minneapolis ninth grade	190	39.44	8.24	.36
3. St. Cloud twelfth grade, 1947	231	39.55	7.50	.52
4. St. Cloud twelfth grade, 1948	271	40.30	7.70	.47
5. University of California advanced graduate students	40	48.58	4.02	.42
II. For the full (76-item) scale				
1. University of California first-year graduate students in psychology	30	63.20	5.12	.44

* The first five samples were scored only for the 59 items taken from the MMPI. The psychology sample took the full scale.

† For the two graduate student samples, the criterion test was the Miller Analogies Test, form g. For the other samples, conventional IQ measures were used.

cially written item would be, "I get sort of annoyed with writers who go out of their way to use strange and unusual words," and of an item from a standard questionnaire, in this case the MMPI [5], "I gossip a little at times."

These items were then reproduced in a booklet and administered to the senior students in four Minnesota high schools.[3] In each school, the responses of students in the upper 25 per cent of the range on a conventional intelligence test were compared with those of students in the lower quartile.[4]

Table 1 provides an example of the way in which the analysis was conducted. Items which showed consistent and significant differences in all Hi-Lo comparisons were retained for the scale. Altogether, 76 items were identified which met this criterion.

These 76 items were scored on the original samples with the results indicated in Table 2. The next step was to cross-validate this scale. Table 3 presents the results observed in six cross-validating samples.[5] In the first five, only MMPI [5] items were available, but the sixth sample took the full 76-item scale. All of the correlations are clearly significant, statistically, and all save one are high enough to be of some practical value. An interesting finding here was the degree to which the scale retained its validity when checked in the graduate student samples where one would encounter rather severe attenuation of the criterion measure.[6]

The individual items were also checked in the two graduate student samples, and it was discovered that 24 of them were contributing only slightly, if at all, to the correlations with

[3]The four communities were: Bemidji, Dawson, Redwood Falls, and St. Cloud. The writer wishes to thank the following persons for making this testing possible: Mr. R. B. Clay, Mr. C. H. Dahlin, Mr. H. B. Gough, Mr. R. Gray, Mr. J. W. Smith, Mr. R. H. Sorenson, Mr. G. Ramseth, and Mr. R. H. Witt.

[4]Because of the small number of cases available, schools A and B were combined into a single sample for this analysis.

[5]The first four samples were made available through the kindness of Mr. M. J. Lunemann, Dr. Starke Hathaway, Mr. F. J. Herda, and Mr. H. B. Gough. The graduate student samples were seen at the University of California Institute of Personality Assessment and Research, Dr. D. W. MacKinnon, Director, as part of a large-scale study of personal and social effectiveness. The Institute projects are being carried out under a grant from the Rockefeller Foundation.

[6]The preliminary 59-item MMPI version was tried in the Michigan clinical psychology assessment program [6]. The abbreviated *Ie* scale yielded significant correlations with the following criterion variables: diagnostic competence (.21); individual psychotherapy (.32); professional interpersonal relations (.28); over-all clinical competence (.23); preference for hiring (.24); and liking (.23).

The correlations for professional interpersonal relations and for individual psychotherapy were both the second highest obtained among all the objective scales utilized in the Michigan study (including the Miller Analogies Test, 6 Primary Mental Abilities test scales, 14 Guilford-Martin scales, 10 conventional MMPI scales, 59 scales on the Strong Vocational Interest Blank, 9 Kuder Preference Inventory scales, and 6 scales from the Allport-Vernon Study of Values).

the criterion tests. It was decided, accordingly, to eliminate them and to reduce the scale to 52 items.[7] This 52-item scale is the one which is currently being used, and it is also the version included in the California Psychological Inventory.[8,9] These 52 items, and the scored responses, are listed below.

1. I have had very peculiar and strange experiences. (False) 2. It often seems that my life has no meaning. (False) 3. When someone does me a wrong, I feel I should pay him back if I can, just for the principle of the thing. (False) 4. Any job is all right with me, so long as it pays well. (False) 5. People often expect too much of me. (False) 6. I get sort of annoyed with writers who go out of their way to use strange and unusual words. (False) 7. Success is a matter of will power. (False)

8. When in a group of people I have trouble thinking of the right things to talk about. (False) 9. It is more important that a father be kind than that he be successful. (True) 10. I was a slow learner in school. (False) 11. I must admit I have no great desire to learn new things. (False) 12. My skin seems to be unusually sensitive to touch. (False) 13. Parents are much too easy on their children nowadays. (False) 14. I read at least ten books a year. (True)

15. I have the wanderlust and am never happy unless I am roaming or traveling about. (False) 16. My parents have often disapproved of my friends. (False) 17. Teachers often expect too much work from the students. (False) 18. I am not afraid of picking up a disease or germs from door knobs. (True) 19. I am quite a fast reader. (True) 20. I have had more than my share of things to worry about. (False) 21. I feel like giving up quickly when things go wrong. (False)

22. In a group of people I would not be embarrassed to be called upon to start a discussion or give an opinion about something I know well. (True) 23. I like to read about history. (True) 24. In school I found it very hard to talk before the class. (False) 25. I dread the thought of an earthquake. (False) 26. I like science. (True) 27. I am bothered by people outside, on street cars, in stores, etc. watching me. (False) 28. I have never been in trouble because of my sex behavior. (True)

29. I often get feelings like crawling, burning, tingling, or "going to sleep" in different parts of my body. (False) 30. I seem to be about as capable and smart as most others around me. (True) 31. I like poetry. (True) 32. I am quite often not in on the gossip and talk of the group I belong to. (False) 33. A windstorm terrifies me. (False) 34. I enjoy a race or game better when I bet on it. (False) 35. Most people make friends because friends are likely to be useful to them. (False)

36. I have never seen a vision. (True) 37. The future seems hopeless to me. (False) 38. I have had no difficulty starting or holding my urine. (True) 39. I often feel as if the world was just passing me by. (False) 40. I get pretty discouraged with the law when a smart lawyer gets a criminal free. (True) 41. The only interesting part of the newspaper is the "funnies." (False) 42. I gossip a little at times. (True)

43. At times I have been so entertained by the cleverness of a crook that I have hoped he would get by with it. (True) 44. I have had attacks in which I could not control my movements or speech, but in which I knew what was going on around me. (False) 45. I have had no difficulty in starting or holding my bowel movement. (True) 46. I like to read about science. (True) 47. I do not read every editorial in the newspaper every day. (True) 48. If people had not had it in for me, I would have been much more successful. (False) 49. I daydream very little. (False)

50. I have often been frightened in the middle of the night. (False) 51. I work under a great deal of tension. (False) 52. I seldom worry about my health. (True)

A review of these items is quite interesting in highlighting some of the personological correlates of general intelligence. The best method of extracting the immediate implications of the items is simply to read the list carefully, but as an aid in forming rational groupings and clusters the following synopsis is provided:

1. Self-confidence and self-assurance, freedom from unsubstantiated fears and apprehensions.

[7] A table listing each item, the percentage in each sample responding "True," and the significance tests has been deposited with the American Documentation Institute. Order document No. 3972 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting $1.25 for 35-mm. microfilm or $1.25 for 6- by 8-in. photocopies.

[8] The California Psychological Inventory, copyright 1951. This inventory is designed to assess a number of social and interactional variables. Its emphasis is not on medical, psychopathological, or maladaptive characteristics, but, rather, on more favorable and positive aspects of individual functioning. It contains 19 scales, including measures for social responsibility, flexibility, academic motivation, tolerance, etc. See, for example, references [3] and [4].

[9] Thirty-nine of the 52 items were taken from the MMPI (by permission of the copyright owners), and the remaining 13 items were original. The 39 MMPI items, and their scorings, are as follows (numbers given are for the MMPI group form booklet): 13F, 28F, 33F, 35F, 36T, 37T, 60T, 62F, 63T, 78T, 116F, 122T, 146F, 194F, 198F, 221T, 224F, 225T, 256F, 260F, 267F, 277T, 280F, 304F, 338F, 392F, 448F, 455F, 462T, 464T, 487F, 492F, 521T, 524T, 526F, 541F, 546T, 552T, 559F.

2. Effective social technique and adjustment, sense of social acceptability without dependence on others; not suspicious, touchy, or overly sensitive.
3. Good physiological functioning, absence of minor, debilitating symptoms and complaints.
4. Liking, and respect, for intellectual pursuits; wide range of interests.

Three additional samples were then obtained in order to cross-validate the 52-item revised scale.[10] Table 4 presents the findings. The validity coefficients are about what they were with the first six cross-validating samples, so it appears that the shortening did not weaken the scale's effectiveness.

Table 4
Summary Statistics for the Revised (52-item) *Ie* Scale in three Cross-validating Samples

Sample	*N*	*M*	*SD*	*r**
1. Rock Island, Illinois high school	461	33.90	7.07	.49
2. Wellsville, New York, high school	230	34.08	6.39	.50
3. University of California undergraduate psychology students	43	40.05	5.16	.42

* Intercorrelation between the Ie scale and a conventional measure of intelligence in each sample.

[10]The Rock Island, Illinois, sample was obtained through the courtesy of Mr. Charles O. Austin, Jr.; the Wellsville sample through the kindness of Dr. H. D. Babcock; and the University sample through the courtesy of Dr. David Krech and Mrs. Ann Miller.

Final high school grade averages were also available for the Rock Island sample. For the females, the correlation between grades and tested IQ was .56, and between grades and the *Ie* scale .59. For males the correlation between grades and IQ was .51 and between grades and the *Ie* scale .31.

In order to provide an over-all picture of the summary statistics at this point, the means and standard deviations for each sample have been proportionately equated to the 52-item test version with the results indicated in Table 5. In this table one can observe a consistent rise in mean scores from junior high school groups, through senior high school and college undergraduate levels, to a final culmination in the graduate student samples.

Adjective Check List

The last data to be presented on this scale have to do with the "social-stimulus values" of persons scoring high and low on it. Forty graduate students seen in an intensive 3-day assessment at the University of California Institute of Personality Assessment and Research were given the scale. Staff judges completed Gough Adjective Check Lists [2] for each as-

Table 5
Ie Scale Summary Statistics for All Samples*

Samples	*N*	*M*	*SD*	*r*	*SE*
I. Original samples					
1. Bemidji twelfth grade	135	35.87	5.82	.47	.09
2. Dawson twelfth grade	35	35.12	6.29	.68	.17
3. Redwood Falls twelfth grade	66	35.16	6.44	.64	.12
4. St. Cloud twelfth grade, 1949	205	35.52	6.50	.54	.07
II. Cross-validating samples					
1. St. Cloud ninth grade	82	31.80	6.52	.53	.11
2. Minneapolis ninth grade	190	34.74	7.26	.36	.07
3. Rock Island high school	461	33.90	7.07	.49	.05
4. St. Cloud twelfth grade, 1947	231	34.84	6.61	.52	.07
5. St. Cloud twelfth grade, 1948	271	35.50	6.78	.47	.06
6. Wellsville high school	230	34.08	6.39	.50	.07
7. University of California undergraduate psychology students	43	40.05	5.16	.42	.15
8. University of California first-year graduate students in psychology	30	43.23	3.50	.44	.19
9. University of California advanced graduate students	40	42.80	3.54	.42	.16

* Scores for samples not taking the final 52-item scale were proportionately corrected to equate them to this standard.

sessee, and then these observer-reports were composited. Adjectives differentiating significantly between the composite reports on the 10 highest and 10 lowest scorers on the *Ie* scale are given below:

1. Adjectives checked more frequently about higher scorers.

adaptable	kind
appreciative	logical
capable	obliging
clear-thinking	persistent
conscientious	planful
considerate	pleasant
cooperative	reasonable
dependable	realistic
energetic	reliable
enterprising	responsible
helpful	sincere
honest	stable
industrious	thorough
interests wide	

2. Adjectives checked more frequently about lower scorers.

aloof	impatient
cynical	individualistic
defensive	moody
deliberate	resentful
effeminate	sarcastic
headstrong	sharp-witted

The general trend of adjectives ascribed is quite clear. High scorers are described as capable, clear-thinking, dependable, logical, realistic, etc., and low scorers as cynical, headstrong, impatient, and resentful.

The adjective findings lend support to the notion that the *Ie* scale is measuring something like personal or intellectual efficiency, with their emphasis on characteristics such as industriousness, planfulness, reliability, and thoroughness.

Summary

A study was made of the relationship between intellectual measures and a series of opinion, attitude, and self-description items, believed, on a priori and theoretical grounds, to covary with intellect. Fifty-two items were found to possess empirical validity. These items were assembled in a scale which yielded median validity coefficients of approximately .58 in four original samples, and of .47 in nine check samples.

Adjectival descriptions of subjects scoring high on the scale tended to stress factors of dependability, intellectual clarity, persistence, and planfulness. The decision was made to call the scale a measure of "intellectual efficiency" in order to reflect these properties.

Received December 3, 1952.

References

1. Altus, W. D., & Thompson, Grace. The Rorschach as a measure of intelligence. *J. consult. Psychol.*, 1949, **13**, 341-347.
2. Gough, H. G. *Predicting success in graduate training.* A progress report. Berkeley: Univer. of California Institute of Personality Assessment and Research, 1950.
3. Gough, H. G. Identifying psychological femininity. *Educ. psychol. Measmt*, 1952, **12**, 427-439.
4. Gough, H. G. What determines the academic achievement of high school students? *J. educ. Res.*, 1953, **46**, 321-331.
5. Hathaway, S. R., & McKinley, J. C. *The Minnesota Multiphasic Personality Inventory.* Minneapolis: Univer. of Minnesota Press, 1943.
6. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

A Repetition of Meyer Williams' Study of Intellectual Control under Stress and Associated Rorschach Factors[1]

Virgil R. Carlson and Richard S. Lazarus
The Johns Hopkins University

For many years psychologists interested in problems of psychological stress have assumed that impairment of performance as a consequence of stress was associated entirely with such personality variables as neuroticism and lack of emotional control. This point of view was supported in a study by Meyer Williams [12], which was aimed at the validation of two indices of "emotional" or "intellectual" control in the Rorschach test. Combining these two indices in the prediction of performance on the Digit-Symbol test under conditions of psychological stress, Williams obtained a remarkably high multiple correlation of .824. This study was important not only because it provided support for certain Rorschach concepts, but also because it related directly to the practical and theoretical problem of predicting behavior under conditions of psychological stress.

There are four principal reasons why this finding of Williams is surprising and requires verification. Some of these considerations have been pointed out in a recent review article on the field of stress [8]. In the first place, it is impossible to define psychological stress in terms of stimulus situations such as threat of failure, electric shock, and similar factors, when the performance of particular individuals is being predicted. Varying motivations and past experience of the subjects must determine the extent to which they are involved in the situation, how they interpret it, and what their response will be.

[1]This research was supported in part by the United States Air Force under Contract No. AF 33(038)-22624 monitored by the Perceptual and Motor Skills Research Laboratory of the Air Training Command Human Resources Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

Second, Williams' stress situation was a combination of stresses, including simultaneously the threat of failure, electric shock, and the presence of observers watching the subjects as they performed. The situation therefore tapped a variety of needs and past experiences which are not likely to be comparable from subject to subject. Any single personality measure such as "emotional control" cannot be expected to correlate highly with responses to so complex a threat situation. Third, certain experiments have recently produced evidence which conflicts with the findings of Williams. For one thing, some of these experiments have shown that "neurotic" or "anxious" subjects may perform better than nondisturbed subjects in certain kinds of stress situations [5,10]. Also, a number of investigators have been unable to predict performance under stress by means of Rorschach measures even when the task used is comparable or the same [1, 4, 6, 9]. Finally, a serious limitation of Williams' study is the fact that it depended upon a sample of 25 cases, which is really too small for a reliable estimate of a correlation.

For these reasons it was important to determine whether Williams' results are repeatable. Since it is possible that his experiment may have contained just the right combination of factors for predicting performance under stress from Rorschach measures, we have made the greatest effort to repeat the study as accurately as possible in every essential detail.

Procedure

Twenty-five white male college students reported individually for two experimental sessions. In the first session the subjects were given the Rorschach test according to the procedure outlined in Beck [2]. After a five-minute

rest period, five 90-second trials were given on the Wechsler-Bellevue Digit-Symbol test [11][2] with a one-minute break between trials. Each subject was told that the Digit-Symbol test is part of a standard test of intelligence, and he was encouraged to do as well as possible.

The second session occurred a day or two later. The subject was first given six more practice trials on the Digit-Symbol test. Again one-minute rest periods intervened between successive trials, and a three-minute rest period was interpolated between the third and fourth trials. The last three of these practice trials were designated the "control period" and constituted an estimate of the subject's nonstress performance. Then the subject was taken immediately to another room for three more Digit-Symbol trials under the stress conditions.

During the "stress period" the subject sat at a desk in a small room behind a one-way screen. He could not see out but could be observed from outside the room. A camera was mounted in front of the screen facing into the room. Three to five persons, including at least one woman, observed the subject's performance.[3] The subject was introduced to the observers beforehand and shown the one-way screen arrangement. Electrodes were strapped to the subject's nonwriting hand and a panel containing a white light and a red light was located on the table in front of him. The experimenters rated each subject, during his performance, on a four-point scale for behavioral signs of stress, such as trembling, sweating, tenseness, etc.

The following instructions were played on a tape recording:[4]

"You are now being observed by a number of psychologists on the other side of the screen who are taking notes and continuous photographs of all your reactions throughout the remainder of this experiment. All directions are to be followed implicitly. Rest your arm attached to the electrodes on the table and keep it there from now on. You will notice that the electrodes on your arm are connected to the panel before you. (White light turned on and left on.) The white light which has just gone on indicates that our shock apparatus has been turned on. You are connected to this apparatus. During the following period you may receive a strong electric shock whenever the observers feel that your test performance is not up to our standards. (Red light turned on.) Whenever the red light goes on, you are not meeting our standards and you are in danger of being shocked—like this! (Shock administered through electrodes, red light turned off.) Now pick up your pencil and look at the test blank before you. It is the same test you took just a few minutes ago. Your instructions are the same as before. At the signal 'Go' work as fast as you can until you are told to stop. Concentrate on your work. Remember, your reactions are being observed and continually photographed and you will be shocked whenever your test performance falls below our standards. Get set for Trial 1. Go!"

Five seconds after the "Go" signal the red light was turned on and left on until five seconds before the end of the trial. Just before the red light went off, the subject received a rather strong electric shock. Ninety seconds after the "Go" signal, the tape recording resumed:

"Stop! Place the test you have just completed on the right side of the table. You will shortly be called upon to repeat the test. Remember, you are being observed and continually photographed and you will be shocked whenever your test performance is below our standards. (Pause) Get set for Trial 2. Go!" This second "Go" signal occurred one minute after the "Stop" signal at the end of Trial 1. The same procedure was repeated for Trials 2 and 3. At the end of Trial 3, the tape recording informed the subject, "The experiment is finished. Remain seated until the electrodes are removed."

The subject was then taken to another room and asked to fill out Winne's Neuroticism Scale [13] and a questionnaire concerning reactions to stressful situations. After completing the questionnaires, he was interviewed in order to determine his subjective reactions to the experimental situation and to explain to him the purpose of the experiment. He was also cautioned to maintain secrecy regarding the experiment. The questionnaire and interview sessions were additions to Williams' procedure.

A control group of 15 subjects was given the same number of Digit-Symbol trials according to the same time schedule. The stress conditions, of course, were not imposed on these subjects.

[2] The test was lengthened sufficiently to preclude the possibility of anyone completing it in the time allowed.

[3] Williams' group of observers consisted of three to five persons, including at least two women. Our one-way screen was approximately 2×3 feet, whereas Williams' screen was 8×4 feet. Our experimental room was also apparently smaller than the one used by Williams.

[4] These instructions were reproduced verbatim from Williams' report.

Table 1
Summary of Rorschach Performance

Rorschach category	Experimental group			Williams' experimental group		
	Mean	Range	*SD*	Mean	Range	*SD*
Sum *C*/Total *C*	.83	.5–1.0	.176	.88	.5–1.2	.20
F + % Total	76.8	50–100	13.31	81.6	70–100	6.59
F + % Color Cards	70.1	0–100	26.40	75.5	50–100	9.28

Results

Each of the Rorschach records was scored both according to Beck [2] and according to Klopfer and Kelley [7], yielding a basis for four quantitative scores. These scores were: (*a*) Sum *C*/Total *C*, (*b*) *F* + % for the total record, (*c*) *F* + % for the color cards alone, and (*d*) Buhler, Buhler, and LeFever's Basic Rorschach Score [3].[5] The first three of these were the Rorschach measures used by Williams.

Sum *C*/Total *C* was determined by summing the weighted numerical values of the various color responses (*FC* = .5, *CF* = 1.0, *C* = 1.5) and dividing the sum by the total number of color responses. A high Sum *C*/Total *C* value represents poor integration of form with color, and supposedly, poor emotional control. *F* + % is the ratio (multiplied by 100) of the number of responses scored *F* + to the number scored *F* + plus the number scored *F* −. A high *F*+% represents good form quality, and, supposedly, a high degree of intellectual control. The Basic Rorschach Score is derived from a diagnostic sign list based primarily on Klopfer's system of scoring. Some 102 separate signs are each given a numerical weight, positive or negative. The algebraic sum of these weights constitutes the Basic Rorschach Score, representing an over-all estimate of the degree of maladjustment or personality disintegration. A high positive Basic Rorschach Score is said to indicate good personality integration.

Table 1 compares the performance of our experimental group with the performance of Williams' group on the pertinent Rorschach variables. The important thing to note here is that the two groups appear to be quite comparable. If anything, our group represents a somewhat greater range of individual differences, at least with respect to the *F* + %.

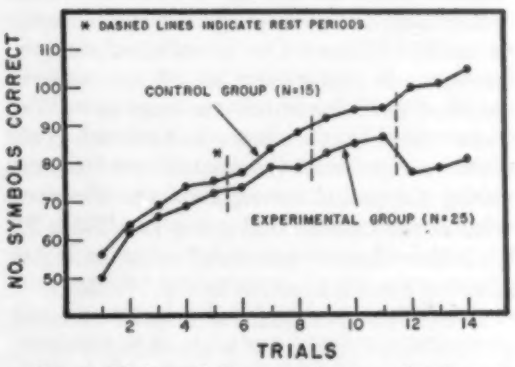


Fig. 1: Digit-symbol test performance for the control and experimental groups.

Comparison of the two groups in performance on the Digit-Symbol test is somewhat more complicated. Williams found that 11 trials were sufficient to reach a plateau of no further improvement. The mean number of digits correctly completed for Trials 9, 10, and 11 (Control Level) minus the number correctly completed for Trials 12, 13, and 14 (Stress Level) was taken as a measure of decrement under stress. Figure 1 shows that our control group certainly, and our experimental group probably, had not reached a plateau by Trial 11. Consequently, Williams' stress score more nearly represents a decrement from the point of maximum performance than does ours. Most likely this accounts for the fact that Williams' average Stress Decrement is nearly twice as large as ours (refer to Table 2). If we were to take the extrapolated learn-

[5] In order to conform with Williams' procedure our Rorschachs were administered according to Beck's instructions, which differ from the procedure set forth by Buhler, Buhler, and LeFever. Our obtained Basic Rorschach Scores may therefore lack strict comparability with those obtained by other investigators.

Table 2
Summary of Digit-Symbol Test Performance

Digit-Symbol measures	Control group		Experimental group	
	Mean	*SD*	Mean	*SD*
(1) Control Level	93.7	18.97	85.1	16.69
(2) Stress Level *	102.1	19.70	79.3	15.19
Stress Decrement (1 minus 2)	−8.4†	4.04	5.8‡ (10.4)§	8.34 (5.70)
t	7.72		3.41	
p	.001		.01	

* A second "control level" for the control group.
† Minus score represents improvement in performance.
‡ If the extrapolated learning curve is used instead of the actual control level, the decrement is 11.0.
§ Values in parentheses are those obtained by Williams.

ing curve of our experimental group as an estimate of the group's maximum performance, we would obtain an average Stress Decrement which is almost exactly equal to the value obtained by Williams. Our actual stress measure, however, is statistically significant and its standard deviation is fully as large as in Williams' data. Further, since our control group shows a significant increase in performance during the period corresponding to the stress trials of the experimental group (see Table 2), we believe that we succeeded in producing as effective a stress situation as did Williams.

In our data the amount of Stress Decrement on the Digit-Symbol test is correlated with the Control Level. This was not the case in Williams' data, and probably the reason is that his subjects had more nearly approached their maximum level of performance. In any event, we derived a set of "corrected" Stress Decrement scores which do not correlate with performance in the control period. We first calculated the predicted Stress Decrement scores by means of the regression equation predicting Stress Decrement from Control Level. Then the obtained Stress Decrement minus the predicted decrement constituted the Corrected Stress Decrement scores.[6] Correlations with the various Rorschach and other personality measures were computed for this measure as well as for the original Stress Decrement. The same result, elimination of the influence of control level of performance, could have been achieved by computing partial correlation coefficients, and any correlation involving Corrected Stress Decrement should be interpreted as a partial correlation.

We found, as did Williams, that subtracting the lowest trial in the stress period from the highest trial in the control period resulted in a set of scores (here called Stress Maximum) which correlated highly with the Stress Decrement (differences between the means of the two periods). Consequently, there was little point in correlating this measure with other variables. Another measure (not used by Williams), however, correlated only slightly with Control Level or with Stress Decrement and so seemed worth while to explore. This is im-

Table 3
Intercorrelations among Digit-Symbol Test Measures

	Stress Decrement	Corrected Stress Decrement	Improvement * under Stress	Stress † Maximum	Stress Level
Control Level	.42 (.05)‡	−.01	.09	.51	.87
Stress Decrement		.90	.33	.94 (.93)	
Corrected Stress Decrement			.32		

* Last stress trial minus first stress trial.
† Highest control trial minus lowest stress trial.
‡ Values in parentheses are those obtained by Williams.

[6] Another method, which is perhaps more logical, consists in predicting stress-level performance from the control level by means of the regression equation predicting Stress Level from Control Level and then subtracting the Stress Level scores from the predicted scores. The scores obtained in this way correlated .996, however, with the Corrected Stress Decrement scores described above, suggesting that it made little difference which method was used.

provement under stress. It was obtained by subtracting the score for the twelfth trial (first trial under stress) from the score for the fourteenth trial (last trial under stress).

The intercorrelations of these various Digit-Symbol measures are presented in Table 3. A correlation of about .40 is required for significance at the 5% level of confidence.[7]

The correlations between the stress measures and the Rorschach measures are shown in Table 4. The two most important correlations in Williams' study were between Stress Decrement on the one hand and Sum *C*/ Total *C* and $F + \%$ on the other. The combined predictive power of these two Rorschach measures was relatively great, yielding a multiple correlation of .824. We do not corroborate this finding at all. Our correlations between Stress Decrement and $F + \%$, and $F + \%$ on the color cards, are near zero. Our correlation between Stress Decrement and Sum *C*/Total *C* approaches significance, but it is in the opposite direction from Williams' correlation. Moreover, this last correlation becomes clearly insignificant when the influence of control performance is removed from the Stress Decrement score. As might be expected, our multiple correlation between Stress Decrement and Sum *C*/Total *C* plus $F + \%$ (.39) is practically the same as the correlation between Stress Decrement and Sum *C*/ Total *C* alone. It might be considered to approach significance except for the fact that it is in the wrong direction. In addition, Stress Decrement failed to correlate with the Basic Rorschach Score, the Behavior Rating, or either of the questionnaires.

Table 4
Correlations between Digit-Symbol Test Measures and Rorschach Measures

	Sum *C* / Total *C*	*F*+% Total	*F*+% Color cards	Basic* Rorschach score
Stress Decrement	−.37 (.35)†	.16 (−.61)	.06 (−.72)	−.17
Corrected Stress Decrement	−.28	.05	.02	−.07
Improvement under Stress	−.07	.12	.03	−.18

* Buhler, Buhler, and LeFever sign check list.
† Values in parentheses are those obtained by Williams.

[7]All correlations referred to in this paper are Pearson product-moment coefficients.

The only positive instance of correlation in all our data was between Improvement under Stress and Winne's Neuroticism Scale. The correlation coefficient in this case was + .54, meaning that the subjects showing greater neuroticism or anxiety on the inventory tended to show more improvement during the stress trials. Also, the correlation between the neuroticism inventory and Behavior Rating[8] approached significance with a coefficient of + .37. Neither the inventory nor the Behavior Rating correlated with any of the Rorschach measures, however.

The most noteworthy fact derived from interviewing the subjects was that they considered the electric shock the most disturbing factor in the situation. Most of them expressed the view that the one-way screen helped them to ignore the presence of the observers. And in general they claimed to have paid little or no attention to the red light, which was supposed to signify below-standard performance. When they did think about the light, it seemed to carry the implication of the imminence of shock rather than failure. It would take some further experimentation to ascertain the actual contribution of each of these factors to the decrement produced in performance. In our own opinion the electric shock did indeed contribute the most, but it is not clear in what precise way. Part of the effect of the shock was apparently just a simple and momentary mechanical interruption in performance.

Discussion

The primary result of the present study is the complete lack of corroboration of Williams' findings. This result, however, agrees with our theoretical expectations concerning the problem of psychological stress and with other experiments having to do with the prediction of performance under stress from Rorschach measures. Experiments by Deese and Lazarus [4], Eriksen, Lazarus, and Strange [6], Stopol [9], and Allee [1] have all produced negative

[8]Each of the experimenters (independently) rated the behavior of the subjects for overt signs of stress during performance on the stress trials. The rating categories ranged from 1 (little indication of stress effect) to 4 (pronounced indications of stress). Since the two independent sets of ratings agreed fairly highly, they were combined into a single Behavior Rating score.

results. In the experiment most similar to Williams' prior to the present report, Stopol found no significant relationship between any Rorschach measure and performance under stress. Since Williams' experiment stands alone, it seems reasonable to reject his results as representative of the relationship between Rorschach variables and performance under stress.

Williams' experiment was undertaken from the point of view of validating certain Rorschach concepts, whereas in the researches cited above the Rorschach measures were secondary to the study of the effects of stress on performance and their prediction. From the viewpoint of Rorschach validity, even if form quality and form-color integration could be properly designated as representing emotional or intellectual control, our theoretical position with respect to the determinants of performance under stress would not lead us to expect a substantial relationship between such "control" and performance under stress. Indeed, one could even make the reverse prediction on the basis of the findings that neurotic individuals may do better than nonneurotics under certain kinds of stress conditions [5, 10].

What is important from an empirical point of view is that these Rorschach measures do not enable us to predict the kinds of behavior which have been studied here. We cannot be certain whether this means that the Rorschach concepts are inadequate or whether the behavior studied is inappropriate to test them. In order to be useful, the concepts of emotional and intellectual control must lead to some valid behavioral prediction.

A final word must be offered with respect to the correlation which was found between Improvement under Stress and the scores on the neuroticism inventory. Although this correlation is large enough to be statistically significant, in a matrix of so many correlations one might obtain such a relationship by chance. But if it can be substantiated, it is undoubtedly related to the finding, mentioned previously, that neurotics sometimes do better than nonneurotics in stress situations.

Summary

An experiment by Meyer Williams, in which a high degree of relationship was found between performance under stress and Rorschach measures of emotional and intellectual control, was repeated with careful attention to every essential detail. The repetition was necessitated by the appearance of conflicting evidence concerning the efficacy of the Rorschach in predicting performance under stress, the belief that a competent theory of psychological stress precluded the finding of substantial relationships, and the fact that Williams' conclusions depended upon a small sample of 25 cases.

The task used was the Wechsler-Bellevue Digit-Symbol test. Each of 25 subjects completed a total of eleven trials on this test under control conditions and then three trials under stress conditions. The stress consisted of electric shock, failure information, and the presence of several people observing the performance of the subject. The difference between the subject's performance during the three trials preceding the stress period and the three stress trials constituted the measure of decrement due to stress. These data were compared with the performance of a control group, indicating stress effects which were quite comparable with those reported in Williams' study.

The performance of our group of subjects on the relevant Rorschach variables was also comparable to that of Williams' subjects with respect to mean, range, and standard deviation. There was no relationship, however, between the Rorschach measures and decrement in performance under stress.

In view of the fact that all other similar investigations have produced negative results, it is reasonable to conclude that the correlation reported by Williams is not representative of the relationship between Rorschach variables and performance under stress.

Received November 17, 1952.

References

1. Allee, Ruth. Rorschach responses of extreme deviation in an experimental stress condition. Unpublished master's thesis, Univer. of Missouri, 1948. Cited in F. McKinney, G. B. Strother, Ruth R. Hines, and Ruth A. Allee. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316-323.
2. Beck, S. J. *Rorschach's test.* I: *Basic processes.* New York: Grune & Stratton, 1944.

3. Buhler, Charlotte, Buhler, K., & LeFever W. D. Development of the basic Rorschach score with manual of directions. Los Angeles: Rorschach Standardization Studies, No. 1, 1948.
4. Deese, J., & Lazarus, R. S. The effects of psychological stress upon perceptual-motor performance. Human Resources Research Center Bulletin, in press.
5. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety-reduction, and stress in learning. Human Resources Research Center Bulletin, in press.
6. Eriksen, C. W., Lazarus, R. S., & Strange, J. R. Psychological stress and its personality correlates. II: The Rorschach test and other personality measures. *J. Pers.*, 1952, **20**, 277-286.
7. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
8. Lazarus, R. S., Deese, J., & Osler, Sonia. The effects of psychological stress upon performance. *Psychol, Bull.*, 1952, **49**, 293-317.
9. Stopol, M. S. An experimental investigation of the consistency of stress tolerance and related Rorschach factors. Unpublished doctor's dissertation, Columbia Univer., 1951.
10. Taylor, Janet A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-93.
11. Wechsler, D. *The measurement of adult intelligence.* (3rd. Ed.) Baltimore: Williams & Wilkins, 1944.
12. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21-29.
13. Winne, J. F. A scale of neuroticism: An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, **7**, 117-122.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Subcategories of Rorschach Human Movement Responses: A Classification System and Some Experimental Results[1]

Robert P. Barrell
Veterans Administration Hospital, New Orleans, Louisiana

Several authors, in Rorschach literature, point to some relationship between various types of human movement responses (*M*) and various personality characteristics, and present clinical hypotheses regarding the interpretation of them. Rorschach [10] himself had some ideas regarding the relationship of extensor and flexor aspects of the movement to activity and passivity within the personality of the subject. Wheeler [13] seems to feel that certain characteristic types of *M* are found in overt homosexuals. Beck [3] suggests that the direction of the movement (away from or towards the center of the card) tells something about the subject's activity or passivity. Piotrowski [9] differentiates several types of movement (along an extensor-flexor scale) as being related to various personality characteristics, relative to self-assertion and submissiveness.

It is common practice in the clinic to utilize the hypotheses referred to above, or similar ones, regarding the meaning of the various types of human movement responses, and to accept them as maxims for interpretation. Yet, aside from the work of Wheeler, there does not seem to be any clear validation of these hypotheses. Thus the problem demands investigation.

The Classification System

It is obvious that for crucial experimental work there must be a clear-cut classification system. This system, it seems, should meet at least two criteria: (*a*) The various categories should be clearly defined, permitting the least possible amount of ambiguity; (*b*) the subcategories should yield a high coefficient of reliability with respect to scoring. Another desirable aspect of such a classification system would be that there be, insofar as possible, a continuum instead of purely categorization, which would allow for more adequate statistical manipulation.

Yet, when the existing classification systems for *M* are examined, not even one is found clearly to meet these criteria. Furthermore, it is obvious that these classification systems overlap each other to a great extent. The Hutt-Milton classification [8] is similar to that of Rorschach [10] in that extensor and flexor aspects are considered; Werner [12] draws closer in his "action," "expression," and "posture" classification; and Zubin [14, 15] cuts across most of these in a different way, considering not only activity but interaction with other people and with the environment in general.

Thus, for purposes of testing some hypotheses regarding the subcategories of *M* [1] it was decided to develop a classification system that would attempt to systematize all of the elements of the existing breakdowns as well as meet the criteria set forth above.

Following Zubin [14, 15] the new classification system of the *M* response was built around two concepts prominent in the previous classifications: (*a*) the amount of *energy* expended in the movement; and (*b*) the amount of *interaction* with the environment shown by the response. Thus was developed, first, the *Energy Classification*, which involved a characterization of the response with respect to the

[1] From doctoral dissertation, University of Michigan, 1950 [1]. The author wishes to acknowledge his indebtedness to the late Professor Louis Granich and to Professors Max L. Hutt, E. Lowell Kelly, Irving H. Anderson, Edward S. Bordin, and Donald G. Marquis.

amount of energy expended in terms of the total body, the limbs and the face, respectively. Next was developed an *Interaction Classification*, which considered: (*a*) degree of interaction with nonanimate objects in the environment; (*b*) degree of interaction with humans specified or implied by the response; and (*c*) the type of interaction with humans specified or implied by the response. Where there was no interaction with humans or with nonhuman objects in the environment, a classification of the *kind* of action involved in the response was used; for example, "a man standing" would be classified as involving no movement in space, and "an angel flying" would show no specific direction of movement.

Further, since it was desired to determine whether sex designation in the *M* response had any relationship to personality variables, a *Sex Designation Classification* was constructed. This consisted of: (*a*) designation as male and how so designated (whether primary or secondary sex characteristics, or both, were referred to, or whether no reason was given for the sex designation); (*b*) designation as female and how so designated; (*c*) absence of sex designation; (*d*) vacillation in sex designation, such as giving one sex in the free association and another in the inquiry, or verbal evidence of confusion concerning the sex of the person seen.

Details of the Classification System

The Energy Classification

1. *Whole body.* (*a*) Human seen as expending energy vigorously with the total body. Examples: fighting, dancing, running, whirling. (*b*) Human seen as expending a modicum of energy. Examples: walking, standing, talking. (*c*) Human seen as expending little energy. Examples: sitting, lying, kneeling; any postural response.

2. *Limbs.* (*a*) *H* or *Hd*,[2] seen as expending energy vigorously with limbs. Examples: hitting, kicking, throwing. (*b*) *H* or *Hd*, seen as expending modicum of energy with the limbs. Examples: shaking hands, playing patty-cake, stirring. (*c*) *H* or *Hd*, seen as expending little energy with the limbs. Examples: pointing, gesturing.

3. *Degree of activity in facial expressions* (regardless of *H* or *Hd*). (*a*) Much energy utilized. Examples: calling for help, shouting. (*b*) Modicum of energy utilized. Examples: laughing, talking, grimacing. (*c*) Little energy utilized. Examples: looking, smoking, peering, looking angry.

[2] *H* is the conventional symbol for *Human*, and *Hd* for *Human Detail*.

The Interaction Classification

1. *Objects in the environment.* (*a*) Human seen as not responding to any object in his environment. Action is entirely self-referred. (*b*) Human attempting to manipulate object or objects in his environment. Examples: pulling, pushing, lifting, carrying. (*c*) Human seen as avoiding object in his environment. Examples: running away from something, cringing. (*d*) Human seen as being manipulated by his environment. Examples: pushed, driven, etc. by some object in his environment (object specified). (*e*) Human just responding to some object in his environment, as holding something. (*f*) Human looking at something in his environment.

2. *Degree of interaction with people.* (*a*) Human not seen as relating to other people in his environment. Action is entirely self-referred. (*b*) Human aware of the presence of other people in his environment, but not taking any direct action as a result of this. Examples: looking at someone, smiling at someone, shielding something. (*c*) Human seen as responding to another human, but amount of action is only moderate. Examples: walking, talking, begging. (*d*) Human involved in much action with other people. Examples: fighting, dancing, working.

3. *Type of interaction with people.* (*a*) Expression of aggression. (*b*) Expression of affection. (*c*) Cooperation with others, as in work activity. (*d*) Subject of aggression. (*e*) Miscellaneous activity, e.g., facing each other. (*f*) Lecturing, advising, talking, gossiping, directing. (*g*) Vacillation, e.g., affection-aggression.

4. *Direction of movement.* (*a*) Movement without any specific direction. Examples: floating in air, keeping time with music. (*b*) Moving up. (*c*) Moving down. (*d*) Moving to the right. (*e*) Moving to the left. (*f*) No movement in space, e.g., standing, lying.

Sex Designation Classification

1. *Human seen as male.* (*a*) No reason given. (*b*) Secondary sex characteristics given, e.g., clothing, shape. (*c*) Primary sex characteristics given. (*d*) Both primary and secondary sex characteristics given.

2. *Human seen as female.* (*a*) No reason given. (*b*) Secondary sex characteristics given, e.g., clothing, shape. (*c*) Primary sex characteristics given. (*d*) Both primary and secondary sex characteristics given.

3. *Vacillation.* Sees male, then female, or vice versa.

4. *No sex designation.* No indication of the sex of the figure seen is presented in the response.

Scoring Procedure

Every *M* response in the subject's Rorschach record is scored for the presence or absence of

each of the subcategories defined and grouped above. Thus the scores on these subcategories used for statistical purposes are the sum of the subject's scores on each subcategory for all of his *M* responses. For example, Subject X's score on the subcategory *Body a* (much expenditure of energy with the total body) would be the sum of all of X's *M* responses in which the person was seen as expending much energy with the total body. Similarly, his score on subcategory *Type a* (expression of aggression) is the sum of all his *M* responses in which the expression of aggression appears. Obviously, the subcategories can be combined into groups and scores can be obtained for the groups, as, for example, the total number of Male designations. Subject X's score here would be, of course, the total number of *M* responses in which he designated the sex as male.

Reliability of Scoring

In order to obtain some measure of the reliability of the scoring, a sample of the records from the study [1] was given to five clinical psychologists whose Rorschach training and experience roughly paralleled that of the author. Prior to the "test scoring" 2 trial records were chosen, 1 being a long one, 1 a short one. Each scorer was trained by the author in the scoring of the above subcategories of *M* on these 2 records, and the results of their scoring here were not included in the reliability study. Subsequently each individual scored 3 different Rorschach records, selected in random fashion. Thus there were available two sets of scores for each of the 15 Rorschach records: one obtained by the author and the other by one of the scorers referred to above. Rank-order coefficients were computed between the author's scoring and the scoring of the other psychologists for each of the *M* subcategories. Table 1 presents the results.

Table 1

Rho Coefficients for Interscorer Reliability

M Subcategory	Rho	*M* Subcategory	Rho
Body *a*	.849	Type *b*	.799
Body *b*	.700	Type *c*	.618
Body *c*	.621	*Type *d*	——
Total Body	.889	Type *e*	.555
*Limbs *a*	——	Type *f*	.691
Limbs *b*	.363	*Type *g*	——
Limbs *c*	.610	Direction *a*	.612
Total Limbs	.524	*Direction *b*	——
Face *a*	.705	*Direction *c*	——
Face *b*	.675	*Direction *d*	——
Face *c*	.597	*Direction *e*	——
Total Face	.627	Direction *f*	.757
Objects *a*	.915	Male *a*	.821
Objects *b*	.453	Male *b*	.923
*Objects *c*	——	*Male *c*	——
*Objects *d*	——	*Male *d*	——
Objects *e*	−.083	Total Male	.863
*Objects *f*	——	Female *a*	.741
People *a*	.875	Female *b*	.403
People *b*	.665	*Female *c*	——
People *c*	.572	*Female *d*	——
People *d*	.284	Total Female	.855
Type *a*	.524	Sex–Vacillation	1.000
		Sex–No designation	.782

* Omitted from study [1] because of insufficient frequency of occurrence in sample to permit statistical manipulation. Hence, omitted here.

When the rho coefficients of Table 1 are converted into r's and the significance is computed, 5 of them fall below the 5% level of significance (rho of .480) and 10 below the 1% level of significance (rho of .60). However, 23 of the 33 relationships yield signficances beyond the 1% level, thus indicating that moderate agreement can be obtained among scorers who have had only a minimum amount of training.

The Experimental Results

The following two generalized hypotheses were constructed for purposes of investigating the relationships between the subcategories of *M* and personality characteristics: (*a*) The various subcategories into which the *M* responses may be subdivided are differentially related to various categories of intellectual functioning; (*b*) various categories of *M* responses show significant relationships to different categories of emotional behavior.

Subjects and Experimental Data

The subjects for our study of the above hypotheses were 121 males who were beginning graduate students in clinical psychology, having graduated from a recognized college or university, and specially selected, from a large group of applicants, for training in the Veterans Administration Training Program for Clinical Psychologists. These subjects had been studied intensively for a period of one week at

Ann Arbor, Michigan, during the summer of 1947, as part of the Veterans Administration Assessment Program for Clinical Psychologists.[3] During this period they had been given many objective, projective, and situational tests, and they were subjected to intensive interviews. There were also available records of scholastic achievement and letters of recommendation. Upon the basis of all of these data, each subject was rated by a staff of experienced clinical psychologists and psychiatrists on a number of intellectual and emotional traits. These ratings are taken as criteria in our study, since we consider them to be reasonably valid measures of what the subjects were like at the time of assessment.

Procedure

1. The 121 Rorschach records were scored for *M* according to procedures outlined by Beck [2] and for the subcategories of *M* according to our new classification system.

2. The frequency distribution for each *M* subcategory was prepared, and the means and standard deviations were computed for each. Fourteen of the 47 subcategories had to be omitted from the study at this point because of insufficient frequency of occurrence in our sample (less than 5 cases per cell) for use of the chi-square technique.

3. Frequency distributions were made, and means and standard deviations were computed for the ratings of the 121 subjects on the various intellectual and emotional traits.

4. The two generalized hypotheses above were restated in terms of the individual subcategory classifications and the individual rating-scale items, the subhypotheses thus constructed being designed to test the generalized hypotheses.

5. The Rorschach variables were divided into two or three categories: (*a*) high and low, with the division as near the mean as possible (approximately 50% in each group); (*b*) high, medium, and low (approximately 33-1/3% in each group). These divisions were made according to the distribution of the scores for each variable. The ratings on the various traits utilized as our criterion variables were similarly divided, but only into high and low categories, with their division as near the mean as possible, dependent upon the shape of the distribution.

6. Chi squares were then computed for the various *M* subcategories and the various rating-scale items, the chi squares computed being chosen to test the hypotheses presented above.

7. Scattergrams were plotted for all of the relationships computed in Step 6 that had a *p* value of less than .20, and the shape of the curve was noted.

8. Pearson *r*'s, or etas, dependent upon the shape of the curve in the scattergram, were computed for all the relationships shown by the chi square to have a *p* value of less than .20.

[3] This program is described in detail by Fiske [4], Kelly [5], and by Kelly and Fiske [6, 7].

Table 2

Level of Significance of the Correlations with Intellectual Variables

M Sub-Category	Broad Interests	Imaginative	Independent Minded	Quality of Intellectual Accomplishments
Body *a*	....	....	.01	.01*
Body *b*	.05	.05	....	.05
Body *c*	....	.01*	.01	.01*
Total Body	.01*	.01*	.01	.01*
Limbs *c*	....	....	....	....
Total Limbs	....	....	....	....
Face *a*	....	....	....	....
Face *b*	....	....	....	.05
Face *c*	....	....	.01	.05
Total Face	....	.05	.01	.01
Object *a*	.01*	.01*†	.01	.01*
People *a*	.01	.01*	.01	.01*
People *b*	....	....	.05	.01*
People *c*	....	.01*†	....	.01
Type *a*	....	....	....	....
Type *b*	....	....	....	....
Type *c*	.01*	....	.05	.01*
Type *e*	....	....	....	....
Type *f*	....	....	....	.05
Direction *a*	....	....	....	....
Direction *f*	.01	.01	.01	.01
Male *a*	.01	.01	.01	.01
Male *b*	....	.05	....	.05
Total Male	....	.01*	.01	.01*
Female *a*	....	....	....	....
Total Female	....	....	....	.01
Sex–Vacillation	....	....	....	....
Sex–No designation	.01	.01*	.05	.01*

* Correlation with R partialed out is significant at the 5% level or better.

† Correlation with M partialed out is significant at the 5% level or better.

This p value was chosen as an arbitrary limit for further exploration since it was believed that when the chi square p values were higher than .20, the probability that a significant correlation existed was very small.

9. To rule out the effect of the total number of responses, R, and the total number of M responses on the results, partial r's, partialing out the effect of R and M, respectively, were computed for all relationships shown by Step 8 to have a correlation significant at the 5% level or greater.

Results

1. *Relationships Between* M *Subcategories and Intellectual Functioning*

Table 2 presents the results relative to the relationships between the M subcategories and intellectual functioning. Here the hypothesis cannot be rejected. Fifty-two of the 112 (zero order) correlations of the subcategories with the intellectual variables are significant at the 5% level or better. Eight of the 28 subcategories are significantly related to "Broad Interests"; 12 of the 28 are significantly related to "Imaginative"; 13 are significantly related to "Independent Minded"; and 19 of the 28 are significantly related to "Quality of Intellectual Accomplishments." Nine of the subcategories are not significantly correlated with any of the intellectual variables.[4]

The subcategories of M, then, are differentially related to the intellectual variables. They tend to correlate with these variables independently of R, but are not independent of R with respect to "Independent Minded." Nineteen of the 52 relationships significant from zero-order correlation remain significant when the effect of R is held constant, but none of these involves a relationship with "Independent Minded." These partial correlation coefficients range from .175 to .601, accounting for from 3 to 36 % of the variance.

It is questionable whether M should be partialed out, as is R, since a score on M inevitably means a score on a subcategory. However, when M is partialed out, we find that the subcategories are independent of M in their predictions in only two cases, which could have occurred by chance.

Since less than 50% of the subcategories of M are independent of R in their predictions of the intellectual variables and only two are independent of M (which could have occurred by chance) in these predictions, it seems that the subcategories are not promising variables for the clinician to rely upon in assessing intellectual abilities.

2. *Relationships between* M *Subcategories and Emotional Variables*

The hypothesis that the various subcategories are differentially related to various emotional variables is not substantiated. Only nine of 132 correlations are significant at the 5% level or better, and only 1 significant at the 1% level or better. This is hardly above chance expectancy, since it would be expected that approximately 7 out of 132 correlations would be significant at the 5% level or better by chance alone. Only 4 of the relationships are independent of R, and 2 independent of M. These correlations could have been obtained by chance.

It may be argued that these low correlations were obtained because of low reliabilities in the criterion traits. However, Kelly and Fiske [7] and Tupes [11] present data on the reliabilities of the ratings. Estimated indexes of reliability of the individual rating-scale items utilized to test the above hypothesis are as follows: .80, .89, .81, .85, .84, .79, .79, .90, .85, .84, .88. Thus it does not appear that the low correlations obtained can be explained on the grounds of low reliability of the criterion data.

It may be that the possible selection of the population on the personality variables does not offer an optimal situation for an adequate test of the hypothesis. Nevertheless, it seems clear that one is not justified, with the type of subjects utilized in this study and with the type of classification of M responses employed, in concluding that the subcategories of M are related to the emotional variables investigated here.[5]

[4] These are: *Limbs c, Total Limbs, Face a, Type a, Type b, Type c, Direction a, Female a,* and *Vacillation in Sex Designation.*

[5] The variables are: "Readiness to Cooperate vs. Obstructiveness"; "Assertive vs. Submissive"; "Suspicious vs. Trustful"; "Generous vs. Self-centered"; "Cautious vs. Adventurous"; "Socially Poised vs. Socially Clumsy"; "Frank vs. Secretive"; "Social Adjustment"; "Appropriateness of Emotional Expression"; "Characteristic Intensity of Inner Emotional Tension"; and "Sexual Adjustment."

Implications

1. M *Subcategories and the Intellectual Variables*

Study of Table 2 shows some interesting relationships and provides for some speculations, to be investigated experimentally at a later date.

First, the *Body* subcategories correlate significantly with the intellectual variables in 12 out of 16 cases. Yet the *Limbs* subcategories do not correlate significantly with *any*, and the *Face* subcategories correlate significantly with only 6 of the 16. It is hypothesized that the *Limbs*, in particular, and also the *Face* subcategories represent something very personal to the subject, have very personal meaning for him. Hence, projection of *M* here is less related to intellect than projection of *M* into the total body, and more related to emotional factors. Along the line of this speculation, it is pertinent to note that in correlations between the subcategories and the emotional variables, the *Body* subcategories correlated significantly with 2 of the 18 variables tested whereas the *Limbs* subcategories correlated significantly in 3 of the 9 relationships tested.

Of greater interest, perhaps, is the following set of observations: the correlation of *Object a* and *People a* (where there is no interaction with objects or people in the environment) with the intellectual variables is significant in all cases, whereas the correlations of the other Interaction classifications are not significant in all cases; the *Cooperation* interaction correlates significantly with three of the intellectual variables, whereas only one other *Type* subcategory (*Lecturing*) correlates significantly with *any* of the intellectual variables and that with only one ("Quality of Intellectual Accomplishments"); the *No Designation of Sex* subcategory correlates significantly with *all* of the intellectual variables, whereas *Vacillation in Sex Designation* correlates with *none*, and designation as *Female* correlates with only one ("Quality of Intellectual Accomplishments"). The hypothesis growing out of these data is that the person who functions well intellectually is the one who isolates (in the Freudian sense) intellect from emotion; that is, emotional conflict, when not isolated, impairs intellectual functioning. The "isolator" would not be concerned with interaction with objects or people, and hence is not particularly concerned regarding the sex of the person seen on the Rorschach nor with highly emotionally charged interaction, projecting only "cooperation."

A third hypothesis growing out of the data is based on the fact that there is more significant correlation between the designation as *Male* and the intellectual variables than there is with the other sex-designation subcategories. The suggestion is that the man who gives much *Male* designation over the other sex-designation subcategories, does so because he has worked out a good masculine identification, and in many ways then can be said to be "better adjusted" than those males who have not worked out a good masculine identification. Hence, he is able to utilize his intellect more effectively.

2. M *Subcategories and the Emotional Variables*

Since the correlations are little above chance, it does not appear that it is possible, at the present time, to utilize the various subcategories of *M* employed here as measures of the emotional traits investigated. Perhaps these subcategories are related to some of the traits not investigated here. Perhaps, also, some combination of these Rorschach variables will be good predictors of these emotional traits. But, at the present stage of investigation, it seems to be a very dubious procedure for the clinician to utilize the subcategories utilized here as measures of the emotional variables tested here. In the event of lack of evidence concerning the relationship between other types of subcategories of *M* and other emotional traits, the negative results here emphasize that the clinician should be doubly cautious about making interpretations from *M* subcategories. It is true, of course, that the clinician may obtain clues to be investigated with other Rorschach data and with other clinical data, as he does with many other clinical hunches. But to assume that there is anything like a one-to-one relationship between the *M* subcategories and emotional variables, or to assume that the *M* subcategories have any clear validity as measures of emotional traits, seems to be an extremely hazardous procedure.

Summary

1. The necessity for experimental verification of hypotheses regarding the interpretation of various types of human movement responses in the Rorschach test is pointed out.

2. A new classification system for *M* responses, embodying elements of existing systems, is developed, in an effort to meet requirements of good experimental work, and presented here.

3. An experimental study of these subcategories of *M* is reported, investigating the following generalized hypotheses: (*a*) The various subcategories into which the *M* responses may be subdivided are differentially related to various categories of intellectual functioning; (*b*) various categories of *M* responses show significant relationships to different categories of emotional behavior.

4. Subjects utilized for this study were 121 males, beginning graduate students in clinical psychology, studied intensively in the Veterans Administration Assessment Program for Clinical Psychologists.

5. Criteria for the study were the ratings of the 121 subjects on a variety of intellectual and emotional traits, made by a staff of experienced clinical psychologists and psychiatrists.

6. Results: Hypothesis (*a*) was substantiated by the data, although it was discovered that the subcategories are not *independently* related to the intellectual variables; hypothesis (*b*) was *not* substantiated by the data.

7. Implications of the findings, particularly as concerned the clinician's use of *M* subcategories in interpretation of the Rorschach, are discussed.

Received December 22, 1952.

References

1. Barrell, R. P. The relationship of various types of movement responses in the Rorschach test to personality trait ratings. Unpublished doctor's dissertation, Univer. of Michigan, 1950.
2. Beck, S. J. *Rorschach's test.* Vol. 1. *Basic processes.* New York: Grune & Stratton, 1944.
3. Beck, S. J. *Rorschach's test.* Vol. 2. *A variety of personality pictures.* New York: Grune & Stratton, 1946.
4. Fiske, D. W. Consistency in the factorial structure in personality ratings from different sources. Unpublished doctor's dissertation, Univer. of Michigan, 1948.
5. Kelly, E. L. Research on the selection of clinical psychologists. *J. clin. Psychol.*, 1947, **3**, 39–42.
6. Kelly, E. L., & Fiske, D. W. The prediction of success in the V.A. training program in clinical psychology. *Amer. Psychologist*, 1950, **5**, 395-406.
7. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.
8. Milton, E. O. The influence of varied experimental sets upon certain Rorschach variables II: Stability of the human movement variable. Unpublished doctor's dissertation, Univer. of Michigan, 1949.
9. Piotrowski, Z. A. Tentative Rorschach formulae for educational and vocational guidance in adolescence. *Rorschach Res. Exch.*, 1943, **7**, 16-27.
10. Rorschach, H. *Psychodiagnostics.* Bern: Huber, 1942.
11. Tupes, E. C. An evaluation of personality-trait ratings obtained by unstructured assessment interviews. *Psychol. Monogr.*, 1950, **64**, No. 11 (Whole No. 317).
12. Werner, H. Perceptual behavior of brain injured mentally defective children: An experimental study by means of the Rorschach technique. *Genet. Psychol. Monogr.*, 1945, **31**, 51-110.
13. Wheeler, W. M. An analysis of Rorschach indices of male homosexuality. *Rorschach Res. Exch.*, 1949, **13**, 97-126.
14. Zubin, J., Chute, E., & Verniar, S. Psychometric scales for scoring Rorschach test responses. *Charact. & Pers.*, 1943, **11**, 277–301.
15. Zubin, J., & Young, Kathleen M. *Manual of projective and cognitive techniques.* Madison, Wisc.: College Typing Co., 1948.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Further Investigation of Rorschach Determinants Subjected to Factor Analysis

Harold L. Williams and James F. Lawrence
Letterman Army Hospital

Utilizing the Thurstone method of centroid analysis, Wittenborn [4, 5] has recently published two studies investigating the patterns of Rorschach scores obtained from two groups of subjects, college students and psychiatric patients. These factor analytic studies were undertaken to evaluate four basic hypotheses [4]: (*a*) the factorial composition for *M* will be different from the factorial composition for determinants involving the use of color; (*b*) the *W, CF, C, c,* and pure diffusion categories are so intercorrelated as to yield a factor differentiated from *F, M,* and *D*; (*c*) there is a productivity factor which (*d*) is involved to varying degrees in the several response categories. In general the hypotheses were confirmed, and his data indicated that *FC,* considered factorially, is more similar to *M* than to *CF* and *C*. His second study revealed some differences between the student and patient groups, introducing a new factor which the author believed was related to mental health.

It is apparent that Wittenborn's results question the adequacy of certain standard Rorschach scoring categories, particularly those of movement, vista, texture, and the color variables. Since these results, if confirmed, have considerable importance for future Rorschach validation, it seemed desirable to conduct a similarly oriented analysis involving a different psychiatric population. The present study, the first unit of a research program involving factorial analysis of the Rorschach, Wechsler-Bellevue, and the Minnesota Multiphasic Personality Inventory, was designed to do this. In addition, in this study, Wechsler-Bellevue Verbal and Performance IQ scores were included to discover whether there are Rorschach determinants which cluster with a standard measure of intelligence to form a separate factor. Rorschach literature [1, 2, 3] has suggested that an intelligence factor should be most heavily loaded with *W, M, R* and accurate form level.

Procedure

Rorschach protocols and Wechsler-Bellevue scores of 86 psychiatric patients were selected at random from the files of the Clinical Psychology Section of Letterman Army Hospital. These protocols were scored in accordance with procedure advocated by Klopfer [1]. The Rorschach scores as well as the verbal and performance intelligence quotients were intercorrelated, using tetrachoric coefficients.[1] The resulting correlation matrix was factored by the Thurstone centroid method. After extraction of the fifth factor the residuals were small and there was no evidence of common factor variance remaining. The centroid factors were then submitted to a series of orthogonal rotations using the criteria for simple structure. The correlation matrix, the centroid factor loadings, and the rotated factor loadings are shown in Tables 1, 2, and 3. The clarity of these factors might be improved by oblique rotation since loadings would then be higher and prediction from the scores improved.

Analysis[2]

Factor I′ appears related to productivity and pure form, containing high loadings on *R, F,*

[1] Tetrachoric *r*'s were used since the short range of many of the Rorschach scoring determinants made the more conventional Pearson *r*'s undesirable.

[2] Because of the magnitude of the standard error of tetrachoric coefficients and the relatively small *N* in the present sample, it was decided to set a lower limit of approximately .40 for a loading to be considered significant.

Table 1

Intercorrelation among 20 Rorschach Determinants and the Wechsler Bellevue Verbal and Performance IQ's for 85 Psychiatric Patients

Symbol	Variable No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
W	1																					
D	2	19																				
d	3	17	80																			
Dd	4	44	55	41																		
S	5	25	25	01	47																	
M	6	37	75	32	67	52																
FM	7	31	46	23	56	00	50															
m	8	39	31	31	64	19	43	58														
k	9	29	29	35	12	04	12	25	25													
K	10	69	24	21	26	15	33	20	60	25												
FK	11	24	22	27	09	38	29	21	20	09	04											
F	12	32	84	80	60	34	47	16	17	34	–02	14										
Fc	13	28	25	10	24	25	26	22	32	32	10	43	00									
c	14	55	48	37	28	–26	25	26	21	30	24	08	46	33								
C'	15	44	19	31	32	–05	10	16	18	16	11	04	03	13	37							
FC	16	43	40	47	58	21	47	50	55	41	30	09	41	40	35	30						
CF	17	55	21	20	40	02	40	29	10	30	33	19	17	37	56	46	40					
C	18	43	09	00	21	–20	13	–16	–05	26	30	–05	12	01	45	31	09	44				
P	19	10	08	23	14	–06	03	45	24	–27	–25	–05	08	25	–01	24	05	–09	17			
R	20	51	92	69	77	34	65	55	40	65	22	41	90	41	59	28	61	49	06	05		
Vb IQ	21	69	25	18	31	00	30	10	00	32	29	04	15	12	44	12	30	22	22	–10	51	
Pf IQ	22	57	48	36	42	31	30	–08	–13	67	35	07	39	20	24	12	30	25	45	15	56	82

Table 2
Centroid Factors

		I	II	III	IV	V	h^2
1	*W*	76	−44	25	−21	−20	91
2	*D*	77	39	−33	28	07	93
3	*d*	61	26	−20	26	31	65
4	*Dd*	76	28	09	03	−31	76
5	*S*	27	31	−11	−41	−16	38
6	*M*	66	27	09	02	−18	55
7	*FM*	51	41	37	06	−16	60
8	*m*	51	33	51	05	−33	74
9	*k*	51	−25	−22	15	09	40
10	*K*	45	−30	26	06	−41	53
11	*FK*	28	22	08	−32	30	33
12	*F*	66	34	−47	28	19	90
13	*Fc*	44	10	24	−36	29	47
14	*c*	60	−27	13	33	29	64
15	*C'*	39	−18	29	10	23	32
16	*FC*	66	14	15	10	−11	50
17	*CF*	55	−23	37	22	18	57
18	*C*	29	−54	25	15	26	52
19	*P*	14	24	22	−23	17	21
20	*R*	97	26	−31	15	10	1.13*
21	*Vb* IQ	57	−53	−29	−23	−24	79
22	*Pf* IQ	69	−45	−53	−33	00	1.08

* Communalities above 1.00 are probably a result of the small *N* in this sample and the large standard error of tetrachoric *r*.

Table 3
Rotated Factor Matrix

		I′	II′	III′	IV′	V′	h^2
1	*W*	−30	82	21	14	29	90
2	*D*	68	47	26	35	24	92
3	*d*	64	33	−02	00	34	64
4	*Dd*	31	53	58	01	20	76
5	*S*	19	24	29	37	−25	38
6	*M*	30	44	47	06	21	54
7	*FM*	22	15	60	14	37	59
8	*m*	05	18	74	07	39	73
9	*k*	18	54	−16	−15	17	40
10	*K*	−31	51	35	−12	20	53
11	*FK*	21	12	04	51	14	33
12	*F*	82	42	01	−11	15	89
13	*Fc*	07	26	08	56	29	47
14	*c*	60	−25	21	06	42	64
15	*C'*	−05	27	−03	09	50	33
16	*FC*	22	45	39	02	34	52
17	*CF*	−05	36	03	−02	66	57
18	*C*	−29	33	−29	−06	49	51
19	*P*	08	−04	15	39	16	21
20	*R*	75	69	18	01	27	1.14
21	*Vb* IQ	−15	86	−08	−05	−15	77
22	*Pf* IQ	10	95	−28	13	−16	1.06

D, *d*, and *c*. The highest loadings in Factor III′ involve movement variables in addition to *Dd* and *FC*. Lack of "perceptual control" [5, p. 471] seems to characterize Factor V′. The finding of this factor in the present sample confirms Wittenborn's impression that there are consistent individual differences in the degree to which subjects are spontaneous or uncontrolled in their perceptual approach.

Few significant loadings are found on Factor IV′, which forces a conservative interpretation. It should be noted, however, that the loadings in the shading variables, *FK* and *Fc*, are consistent with Rorschach scoring procedures, i.e., scoring shading separately from other variables. Factor II′ is of interest in that it supports the clinical impression that the Rorschach can be used as a predictor of intelligence, a finding which has already received support in previous research [2, 3]. It contains variables which have been utilized by Rorschach interpreters to evaluate intelligence, namely, *W*, *M*, *F*, *FC*, and *R*. The appearance of diffuse shading (*K*, *k*) in this factor is perhaps not surprising, since it undoubtedly requires higher than average intelligence to verbalize such percepts.

Comparison

The five factors described above provide additional support for Wittenborn's four hypotheses. Factors I′, III′, and V′ are similar in composition to three of the four factors obtained by that author in his psychiatric sample. Except for *W* and *K*, the fifth factor in this analysis is similar to his "perceptual control" factor. In the present investigation variables *W* and *K* appear on Factor II′, "the intellectual factor." Factor III′ as in his study involves the clustering of *M*, *FM*, and *FC*. However the appearance of *m* in this factor provides some evidence of a separate "movement" dimension. Factor I′ in this sample is similar to the "productivity" dimension found in his psychiatric sample. The presence of *c* in this factor is not clearly anticipated in the previous studies. However, the factorial composition of *c* here obtained is similar to that found in his second study. The factorial pattern of *FC* in the present analysis is also much more similar to that of *M* than to *CF* and *C*. The appearance of *FC* in conjunction with the movement scores further confirms Wittenborn's concept that *FC* is different from the

CF and *C* determinants and is closely allied with *m*, *M*, and *FM* in its behavioral implications.

The shading variables *FK* and *Fc* form a separate factor. Wittenborn did not obtain such a dimension among his scores, and concluded that shading should not be scored and interpreted independently of other variables such as color. Factor IV′, then, provides some support for the common practice of scoring shading separately from other variables, although it does not demonstrate that fine distinctions drawn between the various qualities of shading are warranted. Wittenborn's analysis also generated a factor composed primarily of *S* and *O*, which he suggested was related to level of mental health. Since *O* was not scored in this set of protocols, it is not possible to confirm these findings.

Summary

The present analysis was undertaken for two reasons: (*a*) to replicate an earlier study by Wittenborn on a different population, and (*b*) to discover whether there is an intellectual factor among the several scores in a standard Rorschach scoring system.

Twenty Rorschach determinants and two Wechsler-Bellevue intelligence scores from the protocols of psychiatric patients of Letterman Army Hospital were intercorrelated and the resulting matrix analyzed by the Thurstone centroid method. In the analysis five factors are necessary to account for the common factor variance as compared to four in Wittenborn's two previous analyses. The five factors appear to represent the following dimensions: (*a*) "productivity," (*b*) "movement," (*c*) "lack of perceptual control," (*d*) "shading," and (*e*) "intelligence."

Wittenborn's original four hypotheses were confirmed, as was his hypothesis that *FC* was more closely allied to movement than to color determinants. In addition, the appearance of an "intelligence" factor supports the belief that certain Rorschach determinants covary with intelligence.

Received March 25, 1953.
Early Publication.

References

1. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N.Y.: World Book Co., 1942.
2. Tucker, J. Rorschach human and other movement responses in relation to intelligence. *J. consult. Psychol.*, 1950, **14**, 283–286.
3. Wishner, J. Rorschach intellectual factors in neurotics. *Amer. J. Orthopsychiat.*, 1948, **18**, 265–279.
4. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261–267.
5. Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950, **14**, 469–472.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

The Effect of Test Anxiety on Memory Span in a Group Test Situation

Donald Kaye, Paul Kirschner, and George Mandler
Yale University

The effect of anxiety on test performance has been investigated in several studies in the literature. Generally, this effect has been demonstrated under the condition of the experimental introduction of stress. Williams [6] found the stress effect in the performance of the Digit-Symbol substitution test and Moldawsky and Moldawsky [2] showed a decrement on the Digit Span subtest of the Wechsler-Bellevue scale [5].

In several previous studies [1, 3, 4] an attempt has been made to show the effect of test anxiety level on the individual performance of typical intelligence tests. Briefly, a questionnaire designed to measure specific test anxiety was administered to a large group of subjects. Two ends — Low Anxiety and High Anxiety subjects — of the resulting distribution were then tested. Differential results for the two groups were obtained on the Kohs Block Design test [1], the Digit-Symbol test [4], and the Mathematical Aptitude test [3].

In the study presented here, the effect of test anxiety on memory span in a group test situation was investigated. Moldawsky and Moldawsky [2] have established the effect of situational stress on the Digit Span test in an individual test situation. This study supplements their procedure in examining the diagnostic value of a memory span test in differentiating two groups of *Ss* whose anxiety level has been independently measured.

Procedure

Thirty-six students in introductory psychology courses, who had previously taken the anxiety questionnaire, described in detail elsewhere [3], served as *Ss*. Seventeen *Ss* were taken from the lower 13% of the distribution of anxiety scores and nineteen *Ss* represented the upper 13% of the distribution. None of the *Ss* knew of any connection between the questionnaire and the testing session.

The 36 *Ss* were tested, for purposes of convenience, in three groups of 8, 14, and 14 respectively. At the beginning of a group session the *Ss* were told that the tests they were about to take measured the memory component of the IQ and that their scores would be available afterwards. These instructions were given in line with previous findings [4] that only under these, presumably anxiety-evoking, conditions did a differential effect between the groups appear.

The memory span test lasted twenty minutes and consisted of 28 items: 10 items were number series varying in length from six to ten; 10 items consisted of mixed letters and numbers with the same variation in length and 8 items contained lists of short words (e.g., number, nail, white, foot, light, soot) varying from six to nine words per list. The *Ss* were required to write down as much of an item as they could recall immediately after it was read to them. The correlation between the memory span test and an alternate form was .82 (Pearson r) for the 36 *Ss*.

Results

The tests were scored for the number of correct reproductions. The mean score of the Low Anxiety group was 142.4, for the High Anxiety group 127.9; the t value of 2.37 is significant at the .03 level of confidence.

This result bears out two main points:

1. A memory span test significantly differentiated two groups with differential levels of test anxiety as measured by an anxiety questionnaire. The Low Anxiety group performed significantly better than the High Anxiety group.

2. The differential effect of anxiety levels

operates in group situations as well as individual testing situations when ego-involving instructions are given.

Received October 28, 1952.

References

1. Mandler, G., & Sarason, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
2. Moldawsky, S., & Moldawsky, P. C. Digit span as an anxiety indicator. *J. consult. Psychol.*, 1952, **16**, 115–118.
3. Sarason, S. B., & Mandler, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.
4. Sarason, S. B., Mandler, G., & Craighill, P. C. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.
5. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
6. Williams, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Neurocirculatory Reactions and a Group Stress Situation

George B. Strother and Donald M. Cook [1]
University of Minnesota, Duluth Branch

Abel [1] found that differences in the recall of incompleted and completed tasks by adolescents were related to differences in the Schneider index of autonomic efficiency. Her findings showed that subjects with good autonomic function as indicated by Schneider indices of 11 or higher recalled more incompleted tasks and fewer completed tasks, whereas subjects with low indices (0 - 6) recalled more completed tasks. Stagner [4] interpreted these findings as indicating that the low-index group responds to interruption as a failure situation leading to repressive forgetting, whereas high-index subjects respond to interrupted tasks as a challenge and tend to recall more of them than of the completed tasks.

The purpose of this study was to perform a parallel experiment using an obviously frustrating test situation instead of the Zeigarnik experiment.

The frustrating test situation selected was McKinney's "reporting test" [2] described below. Since it was demonstrated by McKinney that this situation produced marked individual differences in response to the experimental stress, the problem was to determine whether these differences were related to the subjects' Schneider index.

Procedure

The subjects. The subjects were 46 college students enrolled in an introductory course in psychology.

The "index of autonomic efficiency." The Schneider index [3] is designed to determine the efficiency of autonomic regulation of mild cardiovascular stress by assigning point values to pulse rate and blood pressure readings and changes, when reclining and standing, and at intervals following exercise.

The reporting test. The test is divided into three parts: in Part 1 subjects work for 8 minutes without interruption reporting the numbers of each of three different kinds of objects contained in a series of circles. Then for 16 minutes they are compelled to perform the same task against impossible time limits. They are frequently interrupted and told to mark "failure" (*F*) on that portion of their work. Finally, they work again for 8 minutes without interruption. The scores employed in this study were those recommended by McKinney as being most significant—the differences between the percentage correct in Parts I and III, respectively.

Results

Table 1 has been arranged to parallel, as nearly as possible, the summary of Abel's data as presented in Stagner [4, p. 327]. It can readily be seen by comparing the two tables that the results are in striking agreement. The majority of high-Schneider subjects in Abel's

Table 1
Autonomic Function and Response to a Frustrating Situation

Schneider index	Reporting-test reaction, percentage of cases showing: Marked increase in errors (-88 to -9)	Moderate increase in errors (-8 to -1)	No increase or decrease in errors (0 to 10)	*N*
High (11-17)	32	41	27	22
Medium (7-10)	41	41	18	17
Low (0-6)	72	28	0	7

[1]The authors are indebted to Dr. Fred McKinney of the University of Missouri for the use of the "reporting test" materials.

study remember more interrupted tasks; in the present study more high-Schneider subjects respond to the reporting-test situation by showing no loss or actual gain from Condition I to Condition III. Conversely, in Abel's study the majority of low-Schneider subjects remember more completed tasks; in the present study, the majority of low-Schneider subjects show substantial loss under Condition III.

Table 1 was planned to permit direct comparison with Abel's results. As one consequence, the number of cases in the low-Schneider group is small. The present group of subjects, possibly because of age or some other factor, tended to have a higher average than Abel's adolescents. In order to test whether the results so presented did represent a genuine relationship, a product-moment r between reporting-test scores and Schneider indices was computed. Table 2 gives the means, standard deviations, and the intercorrelation of the reporting test and the Schneider index. The coefficient of .30 is significant at the .05 level.

Table 2

Relationship between Schneider Index and the Reporting Test

Variable	Mean	*SD*	*r*	*p*
Reporting test	−8.75	10.09		
			.30	.05
Schneider index	10.46	3.38		

Discussion and Summary

Numerous physiological factors have been suggested as having some relationship to various personality variables. The autonomic nervous system has been the subject of a fair share of such speculation and investigation. It has seemed plausible to assume that such characteristics as emotionality, nervousness, excitability, or lability might be related to individual differences in the excitability of the autonomic nervous system. Much of the theoretical and experimental work dealing with the concept of frustration tolerance seems to assume the existence of a generalized tendency for individuals to react characteristically to a variety of threatening or frustrating situations. Such a generalized tendency might well be at least partially the result of individual differences in autonomic excitability or efficiency. Abel's work seems to be in accord with this hypothesis if we assume that interrupted tasks appear threatening or frustrating to some subjects.

In order to confirm Abel's finding, using a task which is specifically designed to be disturbing to the subjects, these experimenters used McKinney's reporting test on a group of college students whose Schneider indices they had previously determined. The results accord closely with Abel's and indicate that there is an apparently significant relationship between an individual's autonomic reactions and his response to experimental frustration. The Schneider index measures the readiness with which the individual makes a neurocirculatory adaptation to moderate physiological stress; the reporting test measures the individual's ability to adapt behaviorally to psychological stress. There seems to be a measurable degree of relationship between the two.

Received November 25, 1952.

References

1. Abel, Theodora M. Neurocirculatory reactions and the recall of unfinished and completed tasks. *J. Psychol.*, 1938, **6**, 377–383.
2. McKinney, F., Strother, G. B., Hines, Ruth R., & Allee, Ruth A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.
3. Schneider, E. C. *The physiology of muscular activity.* Philadelphia: Saunders, 1933.
4. Stagner, R. *Psychology of personality.* (2nd Ed.) New York: McGraw-Hill, 1948.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Responses of Women to the Thematic Apperception Test

Leonard D. Eron
Yale University

For meaningful interpretations of TAT productions reference should be made to normative data obtained in a standard fashion from well-defined segments of the population. A study by Ritter and Eron [9] has demonstrated that fantasy is important for personality evaluation primarily when it is deviant. This is not to say that fantasy content which is common to a majority of individuals in a given age and cultural group is not important for the understanding of the personality of these individuals. However, in order to diagnose an individual as primarily maladjusted on the basis of projective test responses, account must be taken chiefly of the ways in which responses differ from those of others in the same age and cultural group. It is important that the experimenter not be misled in his interpretations by responses which seem unusual to the "naked eye" but which on more careful investigation are found to be common among subjects of similar age and status.

The results reported here are based on an extension to two groups of female subjects of a methodology for analysis of TAT stories which was developed with some success on male subjects [3, 4]. It is a procedure for noting and summarizing all the important data in a TAT protocol so that it can easily be compared to normative material and significant deviations from expected responses can then be indicated. This can be done both for responses to individual cards by use of what Rosenzweig has termed "apperceptive norms" [10] and for consistencies throughout all 20 cards in a given protocol by using "thematic" norms. Both types of norms are here presented.

Method

Subjects. The subjects for this investigation included two groups of female subjects: Group C, 40 female college students enrolled in a course in introductory psychology at the University of Wisconsin;[1] Group P, 20 women in the third to fifth month of pregnancy and attending the Obstetrical Clinic of the New Haven Hospital.[2] The age range of Group C was 18 to 31 with a median of 23 years. The age range of Group P was 19 to 35 with a median of 24. All the women in group P had IQ scores of at least 100 on the Wechsler-Bellevue Scale; 6 had a college education, 6 high school, and 8 grammar school only; none had ever seen a psychiatrist except in the current investigation, and none was judged by the psychiatrist to have severely debilitating emotional problems beyond what would be expected in a randomly selected normal group. They represented a cultural cross section of the New Haven population. None of the subjects in Group C had ever been or were at that time under psychiatric treatment. Since they were college students, it is assumed their intelligence was at least normal. Culturally they probably represent a more highly selected group than the subjects in Group P, although they did come from all sections of the country.

Procedure. The Thematic Apperception Test was individually administered to all subjects in a standard manner described elsewhere [3]. All 20 cards recommended for adult females by the authors of the test [8] were included.

Analysis of data. Each story was analyzed for emotional tone, outcome, themes, identification of characters and objects, perceptual distortions and unusual formal characteristics.

[1] These subjects were included in an investigation by Dr. Dorothy Terry [11], to whom thanks are expressed for use of these protocols.

[2] These subjects were being studied in connection with research on natural childbirth [6] in which the author participated and which was under the direction of Dr. Lawrence Freedman.

Only one variable was rated or scored at a time and all the stories for a given picture were rated together.

The scales for emotional tone and outcome are the same as those used in previous studies where they were described, their construction was discussed, and their reliability was established [3, 5]. A further check was made on reliability of the scales in this study since different individuals made the ratings.[2] Two-hundred stories, selected randomly from protocols of all the subjects, were rated by two psychologists for emotional tone and outcome. The Pearson *r* between raters for emotional tone was .86 and for outcome was .75. These coefficients are comparable to those established in the original studies and it would seem thus that this is the order of reliability of these scales, regardless of the individuals making the ratings. It was necessary to add an additional category, *x*, to the rating scale for outcome to cover those stories in which no outcome was given, or which were mere descriptions with no action taking place. In the original study there was no provision for such a category since the subjects were always asked to supply an outcome for each story whenever it was omitted.

An attempt was made to use directly the thematic check list devised for the original study with males [2]. It soon became apparent that the identical list could not be used for both sexes. In the first place, eight of the cards recommended for females are different from those used with male subjects, and new themes had to be introduced; also some of the old themes were not applicable. More important than this, though, was the fact that the original list had been drawn up from the point of view of a male hero since it was assumed that the male subjects would identify with the male characters in the story most like himself in age, etc., as suggested by the authors of the test [8]. Logically, this list could not then be used with female subjects if it is assumed that in general they would in the same way identify with female characters. However, for many of the cards it seemed that the central characters in the stories told by a majority of the female subjects were males. This was especially true for cards 4 and 13 MF, both of which contain a male and a female figure. To check on this, two female experimenters went through all the responses to these cards and made judgments as to which character in each story seemed to be the central one from whose point of view the story took shape. Percentage agreement as to choice of central character was 94 and each examiner selected the female character as central in 26% and 28% of the stories for card 4, and in 4% and 6% of the stories for card 13MF. This result is similar to that obtained in an unpublished study by Cohen [1] in which 50 women subjects were asked who the central character was in their stories. For card 4 he found that 80% of these female subjects chose the male, 20% the female, and for card 13MF, 86% chose the male and 14% the female. Thus it seemed that with female subjects a check list had to be used which was not drawn up from the point of view of a hero or heroine, at least insofar as the heterosexual situation was concerned. Consequently the old check list was revised and Table 1 sets forth this new check list for thematic analysis.

Table 1

Check List of Themes for Analysis of Women's TAT Stories

I. Disequilibrium (tension)

A. *Interpersonal*

1. Parent

a. ***Pressure***—parent or parent figures are prohibitive, compelling, censuring, punishing, disapproving, interfering, checking up, disagreeing with, quarreling with, restraining or unduly influencing child.

aa. Child is exerting pressure on parents.

b. ***Succorance***—child seeks or receives aid, help, advice, consolation from parent.

c. ***Nurturance***—child bestows or offers advice, aid, consolation to the parent.

d. ***Aggression from***—physical harm inflicted or intended upon child by parent.

e. ***Aggression to***—physical harm inflicted or intended upon parent by child.

f. ***Departure***—child is taking leave of parental home, runs away from home.

g. ***Concern***—parent is worried over physical or mental well-being of child.

[2] Mrs. Ruth Straatsma rated all the stories for emotional tone, and Mrs. Margaret Ingram rated those stories used in the reliability estimate. Mrs. Sophie Dutton rated all the stories for outcome, and reliability was checked with ratings made by the author. The author analyzed all stories for themes, unusual formal characteristics, and perceptual distortions.

gg. *Concern*—child is concerned over parent.
h. *Incest*—actual or contemplated.
i. *Death or illness* of parent.
j. *Death or illness* of child.
k. *Disappointment to*—parent is disappointed in child's behavior or accomplishments. Parent rejects child.
l. *Disappointment in*—child is disappointed by parent, ashamed of parents, derogates parent, rejects parent.
m. *Filial obligation*—child feels it is his duty to remain with, comply with, or support parents.
n. Confession—child tells parent of some misdemeanor or crime he has committed.
nn. *Confession*—parent confesses to child.
o. *Bad news*—child brings sad tidings to parent (e.g., death of other parent, brother, etc.).
oo. *Bad news*—parent brings sad tidings to child.
p. *Marriage*—child tells parent of past or impending marriage, parent objects to marriage, interferes with marriage.
q. *Collusion*—parent and child are planning or executing together some antisocial act.
r. *Parental conflict*—child is concerned over marital problems of parents. Parents disagree over child care, etc.
s. *Lost*—child has been abandoned by parents or is lost.
w. Facts of life explained to child by parent.
ww. Parent tells child a sibling is coming.
x. Parent is jealous of child.

2. Partner (husband, sweetheart, opposite sex peer)
a. *Pressure*—heroine and partner are quarreling, can't agree.
b. *Nurturance*—male character bestows or offers to partner aid, consolation, advice, etc., or these are being sought by female.
c. *Succorance*—male character seeks or receives aid, comfort, consolation, assistance from partner, advice from partner.
d. *Aggression from female character*—physical harm inflicted upon or intended for partner by female.
e. *Aggression from male character*—physical harm inflicted upon or intended for partner by male.
f. *Departure from*—male character leaves partner either temporarily or permanently.
ff. *Departure from*—female character leaves partner either temporarily or permanently.
g. *Concern*—male character is worried over physical or mental well-being of partner.
gg. *Concern*—female character is worried over physical or mental well-being of partner.
h. *Illicit sex*—extra- or premarital heterosexual relationship, nonincestuous, includes "petting."
i. *Illicit sex with violence*—rape.
j. *Death or illness* of female partner.
jj. *Death or illness* of male partner.
k. *Disappointment*—male character is disappointed in female's behavior, or accomplishments, rejects female, derogates female.
l. *Disappointment*—female character is disappointed in male's behavior, or accomplishments, derogates male, rejects male.
m. *Jealousy*—female character is jealous of partner's attention to others or other's attentions to partner.
n. *Competition*—male character is jealous of partner's attention to others or other's attention to partner.
o. *Unfaithful*—male character discovers partner has been having extramarital relations or finds her raped.
oo. *Unfaithful*—female character discovers partner has been unfaithful.
p. *Decision*—male character must choose between marriage and not or between two partners.
pp. *Decision*—female character must choose between marriage and not or between two partners.
q. *Pursuit*—male character is wooing or trying to get partner to submit, wants contact with other sex.
r. *Seduction*—female character is actively pursuing or talking male into relationship, wants contact with other sex.
s. *Unrequited*—male character's love is unreturned, impotence. Waits for partner.
ss. *Unrequited*—female character's love is unreturned, is "done in" by partner. Waits for partner.
t. *Restraint*—central character and partner can't be married or have intercourse (because of economics, deformities, menstrual periods, etc.).
u. *Childbirth*—pregnancy.
y. *Surprise*—female character surprised by partner.

3. Peer
a. *Pressure*—like-sexed peers are prohibitive, compelling, censuring, punishing, disapproving, interfering, restraining, etc.
b. *Succorance*—central character seeks or receives aid, advice, consolation from peer.
c. *Nurturance*—central character bestows or offers aid, advice, consolation upon peer.
d. *Aggression from*—physical harm inflicted upon or intended for central character by peer.
e. *Aggression to*—physical harm inflicted upon or intended for peer by central character.
f. *Desertion*—central character has been deserted by his companions or is lost from them.
g. *Concern*—central character is worried about physical or mental well-being of friends.
h. *Homosexuality*—actual or contemplated.

i. *Death or illness* of friend.
j. *Belongingness*—desire expressed to be with or accepted by peers.
k. *Unappreciated by peers.* Rejected by peers, peers not interested in welfare.
l. *Competition*—central character is competing with peer in game or contest, or professionally.
m. *Hypnotism.*
n. *Envy*—central character is envious of peer's accomplishments, capacities, belongings, etc.
nn. Others envy central character.
o. *Revenge*—central character is anxious to extract revenge from a peer for some past deed or is taking the revenge.
p. *Disregard for*—central character is unconcerned over welfare of peers, deserts peers.
q. *Devaluation of peer*—derogation, rejection of peer by central character.

4. Sibling
a. *Pressure*—siblings are prohibitive, compelling, censuring, punishing, disapproving, interfering, restraining, etc.
b. *Succorance*—central character seeks or receives aid, advice, consolation from sibling.
c. *Nurturance*—central character bestows or offers aid, advice, consolation upon sibling.
d. *Aggression from*—physical harm inflicted upon or intended for central character by sibling.
e. *Aggression to*—physical harm inflicted upon or intended for sibling by central character.
f. *Rivalry*—siblings are striving for same end, are jealous of each other, etc.
g. *Concern*—central character is worried about physical or mental well-being of sibling.
h. *Incest*—actual or contemplated.
i. *Death or illness* of sibling.

B. *Intrapersonal*

1. *Aspiration*—dreaming of future, hoping for future, determination.
2. *Inadequacy*—realization, whether justified or not, of lack of success; individual is at a loss to cope with situation.
a. *Unrealized.*
3. *Curiosity*—wondering about construction of object, contents of room, etc., desire to observe, inquire, explore, investigate; to acquire facts.
4. *Behavior disorder*—personal maladjustment of all sorts, neurosis, psychosis, sleeplessness, hallucination, withdrawal, eccentricity, morbid preoccupation, nightmare, excessive emotionality.
5. *Suicide*—attempted or complete, preoccupation with.
6. *Moral struggle*—concern over what is right and wrong, hesitancy in indulging in some act because of ethical proscriptions.
7. *Guilt*—remorse, shame.
8. *Drunkenness.*
9. *Fear-worry*—central character is apprehensive, alarmed, terrified, of some person, thing, force.
10. *Rumination*—inability to understand world and its problems, adolescent reverie, man's insignificance, "what am I?", "what's it all about?".
11. *Occupational concern*—deciding between jobs, considering vocations, dissatisfied with present employment or present pastime; worried about job. Emphasis on decision which must be made in choosing occupation.
12. *Physical illness or death of central character*—(other than when in a peer, partner, sibling, parent relationship, which are included under those specific categories).
13. *Retribution*—forced to atone or be punished for some antisocial act.
14. *Reminiscence, sad*—individual is unhappy in his memories of the past or contemplation of the future, hopelessness.
15. *Intra-aggression*—does physical harm to self short of suicide, (which is carried under that heading), either through accident or deliberate, looks forward to own death.
16. *Religion*—prayer, seeking consolation from God, religious conflict, religious awakening.
17. *Loneliness*—central character misses someone, is an outcast, friendless, homeless.
18. *Compensation*—when individual has one characteristic or stroke of fortune to make up for another bad characteristic or misfortune.
19. *Vacillation*—wasting time, putting off a distasteful task, procrastination, loitering, waiting for someone unspecified.
20. *Acquisition*—desire expressed by central character to acquire material things, or is working greedily for possession of goods; miserliness.
21. *Exhaustion*—characters completely "pooped" from overexertion.
22. *Revenge*—central character is preoccupied with wish to retaliate for some past wrong from unspecified individual (when peer, carried under that heading).
23. *Inconsequential sadness*—over death of dog, broken toy, loss of object, etc.
24. *"Hurt Feelings"*—central character is very sensitive over some happening, slight, insult, etc.
25. *Jealousy*—of unspecified individual (if specific person named, carried under appropriate heading).
26. *Self-pity*—central character feels sorry for self.
27. *Supernatural*—return of the dead, ghosts.
28. *Homesickness*—expressed by central character.
29. *Physical illness or death* of unspecified character (not central character, partner, peer, sibling).
30. *Bored*—central character is bored, has nothing to do.
31. *Regression*—central character is unwilling to grow up, assume adult responsibilities, grow old.
32. *Posing*—central character posing for portrait.
33. *Accusation*—central character accused, usually falsely, of some misdemeanor, crime, infidelity.

C. *Impersonal*

1. *Economic pressure*—individual is compelled to, prohibited from, or limited in doing something because of the lack of money.
2. *Legal restriction*—individual is incarcerated, arrested, or detained against his will.
3. *Generalized restriction*—environment is generally frustrating because of backwardness, danger, lack of opportunity, etc.; hopelessness of life.
4. *Aggression towards environment*—robbery, accident, murder of unspecified individual (if individual named, carried under the proper heading).
5. *Aggression from*—impersonal source, accident, animal, nature, disease, etc.
6. *War*
7. *Escape from perilous (restricting, shameful) environment*—individual is in the act of getting out (doesn't include rescue, which comes under II).
8. *Obligation to society*
9. *Crime without punishment*

II. Equilibrium

A. *Interpersonal*

1. Parent
   - *a. Cooperation*—parent is working with child towards his own goal.
   - *b. Resignation*—parent is resigned to child's activity.
   - *c. Idealization*—child idealizes parent, wants to be like parent.
   - *d. Reunion*—of child with parents.
   - *e. Fulfillment*—child lives up to expectations of parents.
   - *f. Contentment*—blissful home situation, stress security of home.
   - *g. Ordinary familial activity*—just talking, spending an evening, working, etc.
2. Partner
   - *a. Admiration*
   - *b. Cooperation*—between mates, so that both are working for the same end and are happy.
   - *c. Contentment*—mutual love and affection.
   - *d. Reunion*
   - *e. Ordinary activity*
   - *f. Proposal*
   - *g. Anniversary*
   - *h. Dancing*
3. Peer
   - *a. Cooperation*—working together for same goal.
   - *b. Congeniality*—sociability, friendliness, satisfactory affiliation.
   - *c. Reunion*—with friends.
   - *d. Approbation*—central character is lauded, appreciated by friends, and peers.
   - *e. Exhibition*—central character is amusing, entertaining, attracting others.
   - *f. Ordinary activity*—everyday association with nothing in particular happening.
4. Sibling (no II themes found in original sample of 200 cases).

B. *Intrapersonal*

1. *Self-esteem*—confidence, belief in own superiority, conceit.
2. *Tranquility*—peace of mind, content with environment and own accomplishments, enjoyment, aesthetic appreciation.
3. *Reminiscence, happy*—individual is content with her memories, or contemplation of future.
4. Retirement—central character is asleep, resting, no physical exhaustion, relaxation.
5. *Occupational satisfaction.*
6. *Resignation to lot*—individual accepts situation without further struggle.
7. *Ordinary activity*—individual is going about her everyday activities.

C. *Impersonal*

1. *Favorable environment*—individual is being helped by favorable circumstances, is enjoying his surroundings.
2. *Rescue*—successful escape from perilous environment or cessation of noxious stimuli (already accomplished).

The list of unusual formal characteristics is the same as that used in the former studies [3, 9] under the heading of levels of interpretation and includes such things as autobiographical stories, symbolic or abstract stories, descriptions rather than narratives, denial of popular themes, rejections and alternate plots. Perceptual distortions pertain to any marked deviations from the descriptions of objects and characters furnished by the authors in the manual of instruction [8].

Results

Analysis in these terms (emotional tone, outcome, themes, formal characteristics, and perceptual distortions) are presented for stories in response to each card and also for all 20 stories in a protocol, so that there is information concerning what are the common apperceptive responses for each card and also an indication of how many individuals have a given number or more of each kind of response in a total of 20 stories. In the apperceptive norms, the data are presented in terms of percentage of subjects having a specific response for any one card, and in the thematic norms, in terms of the percentage of subjects having a given number or more of these responses in a total of 20 stories. Thus, by referring to Table 2, for example, it can be

seen that 52% of the subjects gave stories with emotional tone rating of −1 to card 1; and by referring to Table 7, it becomes apparent that 75% of the subjects contributed seven or more stories with emotional tone rating of −1 in a total series of 20 stories. Significant differences beyond the .05 level of confidence between the two samples of subjects were noted. In calculating these differences the actual frequencies rather than percentage frequencies were used in the standard formula for calculating the chi square of homogeneity, with Yates's correction where appropriate [7]. Fourfold tables, denoting presence or absence for each group, were utilized in calculating differences in themes, unusual formal characteristics and perceptual distortions. With emotional tone and outcome, where 5 × 2 tables were possible because of rating-scale intervals, appropriate cut-off points were used so that there would be a theoretical frequency of at least five in each cell. Consequently there were a number of 3×2 as well as 2 × 2 tables. In the thematic norms, which indicate the percentage of individuals having a given number or more of any response in a total of 20 stories, the chi squares were calculated from the actual frequency of individuals having each number of instances in question. Cut-off points here were also selected so that the theoretical frequency of any cell would be at least five and there were thus a number of 3 × 2 as well as 2 × 2 tables.

Table 2

Percentage Frequency of Subjects Contributing Stories of Various Emotional Tone Rating for Each of 20 Cards

Card	*Emotional Tone Ratings*					
	−2	−1	0	+1	+2	Reject
1	0	52	22	22	5	0
2	0	48	37	15	0	0
3GF	38	62	0	0	0	0
4	12	87	0	0	0	0
5	3	45	48	3	0	0
6GF	47	28	25	0	0	0
7GF	5	23	50	22	0	0
8GF	22	62	17	0	0	0
9GF	13	72	12	2	0	2
10	13	27	3	43	13	0
11	13	57	30	0	0	0
12F	27	23	37	13	0	0
13MF	77	22	2	0	0	0
14	3	13	40	42	2	0
15	47	42	12	0	0	0
16	2	13	35	37	10	3
17GF	12	27	43	17	0	2
18GF	43	52	3	2	0	0
19	5	27	27	40	0	2
20	7	40	52	0	0	2

Table 3

Percentage Frequency of Subjects Contributing Stories of Various Outcome Ratings for Each of 20 Cards

Card	Outcome Ratings						
	−2	−1	0	+1	+2	?	x
1	3	25	7	28	18	7	12
2	5	17	17	42	12	5	2
3GF	22	27	5	30	5	5	7
4	3	27	2	35	10	10	13
5	7	15	18	25	7	2	25
6GF	5	15	13	33	5	3	25
7GF	3	13	13	37	8	0	25
8GF	8	2	25	25	8	2	30
9GF	10	22	12	27	3	0	27
10	3	17	8	13	37	2	20
11	18	5	12	5	30	2	28
12F	7	23	10	23	10	3	23
13MF	47	8	3	13	0	8	20
14	2	13	13	40	17	2	13
15	23	3	15	27	3	5	23
16	2	8	15	37	17	0	22
17GF	12	15	18	22	7	7	20
18GF	30	15	0	32	3	5	15
19	3	7	8	27	8	0	47
20	10	10	18	23	12	3	23

Table 4

All Themes Contributed by Two or More Subjects for Each of 20 Cards

Code	Theme	Percentage frequency
Card 1		
IA1*a*	parental pressure	60
IB11	occupational concern	37
IB1	aspiration	35
IA3*j*	belongingness	27
IB2	inadequacy	10
IB7	guilt-remorse	5
IC1	economic pressure	5
IB3	curiosity	3
IB19	vacillation	3
Card 2		
IB1	aspiration	52
IB11	occupational concern	27
IA1*a*	parental pressure	23
IA1*m*	filial obligation	15
IIA1*a*	cooperation with parent	10
IC3	generalized restrictions	8
IA4*f*	sibling rivalry	8
IA1*c*	nurturance to parent	8

IA2*r*	seduction	7
IIA1*g*	ordinary familial activity	7
IC1	economic pressure	7
IB7	guilt	5
IA1*f*	departure from parent	3
IA1*i*	death or illness of parent	3
IIA1*e*	fulfillment, parent	3

Card 3GF

IA1*j*	death or illness of child	15
IA2*a*	pressure from partner	15
IA2*ss*	unrequited female partner	13
IA1*a*	parental pressure	12
IB7	guilt	12
IB29	illness or death of unspecified character	12
IA2*f*	departure from partner	10
IA2*jj*	death or illness of male partner	10
IB17	loneliness	8
IA1*i*	death or illness of parent	8
IB12	physical illness or death of central character	7
IB24	hurt feelings	5
IA1*r*	parental conflict	3
IA2*oo*	unfaithful male partner	3
IB2	inadequacy	3
IB4	behavior disorder	3
IC1	economic pressure	3

Card 4

IA2*c*	succorance from partner	38
IA2*f*	departure from partner	25
IA2*b*	nurturance to partner	18
IA2*ss*	unrequited female partner	13
IA2*p*	decision for male partner	12
IA2*r*	seduction	10
IA2*oo*	unfaithful male partner	8
IB2	inadequacy	7
IB11	occupational concern	7
IC5	aggression from environment	7
IA2*gg*	concern of female partner	5
IA2*h*	illicit sex	5
IB1	aspiration	5
IA2*m*	jealousy	3
IA2*n*	competition	3
IA2*q*	pursuit of partner	3
IA2*s*	unrequited male partner	3
IA3*e*	aggression to peer	3
IB6	moral struggle	3
IB22	revenge	3
IC6	war	3

Card 5

IB3	curiosity	48
IA1*a*	parental pressure	35
IA1*g*	concern of parent	17
IB9	fear, worry	12
IIB7	ordinary activity	8
IA1*k*	disappointment to parent	7
IA2*gg*	concern of female partner	5
IC4	aggression to environment	5
IB4	behavior disorder	5

Card 6GF

IA2*a*	pressure from partner	28
IB9	fear, worry	17
IIA2*e*	ordinary activity	15
IA2*q*	pursuit of partner	10
IA2*r*	seduction	10
IA1*a*	parental pressure	7
IB33	accusation	7
IB7	guilt	5
IC4	aggression to environment	5
IA1*p*	marriage	3
IA2*h*	illicit sex	3
IA2*l*	disappointment in male partner	3
IA2*oo*	unfaithful male partner	3
IA2*y*	surprise by partner	3
IB24	hurt feelings	3
IIA2*d*	reunion with partner	3
IA1*g*	concern of parent	3

Card 7GF

IA1*a*	parental pressure	32
IA1*v*	facts of life	18
IA1*w*	a sibling is coming	12
IA3*j*	belongingness	10
IA4*f*	sibling rivalry	10
IB1	aspiration	8
IB7	guilt	8
IB30	bored	8
IB23	inconsequential sadness	7
IIA1*g*	ordinary familial activity	7
IA1*b*	succorance from parent	5
IA*ll*	disappointment in parent	5
IIA1*f*	contentment with parent	5
IB3	curiosity	3
IB4	behavior disorder	3
IB11	occupational concern	3
IB16	religion	3

Card 8GF

IIB3	rèminiscence, happy	33
IB1	aspiration	20
IB11	occupational concern	15
IC1	economic pressure	15
IIB7	ordinary activity	15
IB32	posing	13
IB21	exhaustion	10
IB14	reminiscence	8
IB10	rumination	7
IIB4	retirement	5
IA1*b*	succorance from parent	3
IA1*g*	concern for child	3
IB13	retribution	3
IB17	loneliness	3
IC3	generalized restriction	3
IIB2	tranquility	3

Card 9GF

IC7	escape from perilous environment	33
IB3	curiosity	28
IA2*m*	jealousy	22
IB9	fear, worry	18

IA3*a*	pressure from peer	13
IA2*oo*	unfaithful male partner	13
IA4*f*	sibling rivalry	10
IA2*a*	pressure from partner	7
IA2*e*	aggression from male partner	7
IA2*ss*	unrequited female partner	7
IA3*b*	succorance from peer	7
IC5	aggression from environment	7
IIA3*b*	congeniality with peer	7
IA3*n*	envy of peer	5
IA3*q*	devaluation of peer	5
IA4*a*	pressure from sibling	5
IA2*jj*	death or illness of male partner	3
IA3*c*	nurturance to peer	3
IA3*k*	unappreciated by peer	3
IA3*p*	disregard for peer	3
IB2	inadequacy	3
IB4	behavior disorder	3
IB5	suicide	3

Card 10

IIA2*c*	contentment	40
IA1*j*	death, illness of child	20
IA2*b*	nurturance to partner	18
IA2*c*	succorance from partner	17
IC6	war	13
IA1*g*	concern for child	10
IIA2*h*	dancing	10
IIA1*d*	reunion with parents	8
IA2*f*	departure from partner	7
IC5	aggression from environment	7
IIA2*d*	reunion with partner	7
IIB3	reminiscence, happy	7
IIA2*g*	anniversary	5
IA1*c*	nurturance to parent	3
IA1*f*	departure from parent	3
IA2*t*	restraint from partner	3
IIA1*b*	resignation to child	3

Card 11

IC5	aggression from environment	38
IC7	escape from perilous environment	18
IB9	fear, worry	15
IA3*e*	aggression to peer	13
IB3	curiosity	10
IB12	illness or death of central character	8
IA3*a*	pressure from peer	7
IA3*d*	aggression from peer	5
IB4	behavior disorder	5
IC3	generalized restriction	5
IC4	aggression to environment	5
IC1	economic pressure	3
IC6	war	3

Card 12F

IA1*l*	disappointment in parent	22
IA1*a*	parental pressure	20
IA1*b*	succorance from parent	10
IB14	reminiscence, sad	10
IB4	behavior disorder	8
IB7	guilt	8
IB31	regression	8
IB32	posing	8
IA1*c*	nurturance to parent	7
IB9	fear, worry	7
IA1*i*	death, illness of parent	5
IA1*k*	disappointment to parent	5
IA1*m*	filial obligation	5
IC1	economic pressure	5
IIB1	self-esteem	5
IA1*p*	marriage	3
IA1*q*	collusion with parent	3
IA2*m*	jealousy of partner	3
IA3*q*	devaluation of peer	3
IB2	inadequacy	3
IC5	aggression from environment	3
IIB3	reminiscence, happy	3

Card 13MF

IA2*j*	death, illness of female partner	78
IB7	guilt	42
IA2*e*	aggression from male partner	15
IA2*a*	pressure from partner	13
IC1	economic pressure	13
IB5	suicide	12
IA2*h*	illicit sex	10
IA2*i*	illicit sex with violence	7
IA2*o*	unfaithful male partner	7
IA2*u*	childbirth	7
IB2	inadequacy	7
IB4	behavior disorder	7
IB11	occupational concern	7
IC5	aggression from environment	7
IA1*j*	death, illness child	5
IB13	retribution	5
IA2*s*	unrequited male partner	3
IB6	moral struggle	3
IB8	drunkenness	3
IB16	religion	3
IB21	exhaustion	3

Card 14

IB1	aspiration	32
IIB7	ordinary activity	10
IB11	occupational concern	8
IIB2	tranquility	8
IC1	economic pressure	7
IIB3	reminiscence, happy	7
IA3*j*	belongingness	7
IA1*a*	parental pressure	5
IB4	behavior disorder	5
IB5	suicide	5
IB10	rumination	5
IB17	loneliness	5
IIC1	favorable environment	5
IA2*s*	unrequited male partner	3
IA2*t*	restraint from partner	3
IB2	inadequacy	3
IC3	generalized restriction	3
IC6	war	3

Card 15

IB15	intra-aggression	30
IB16	religion	30
IA2*j*	death, illness female partner	22
IA3*i*	death, illness friend	20
IB4	behavior disorder	18
IB17	loneliness	18
IB7	guilt	15
IB20	acquisition	15
IB14	reminiscence, sad	12
IB27	supernatural	12
IC4	aggression to environment	8
IA1*j*	death, illness child	8
IA3*e*	aggression to peer	5
IA3*k*	unappreciated by peer	5
IA3*p*	disregard for peer	5
IIB7	ordinary activity	5
IB29	illness or death of unspecified character	3
IA3*j*	belongingness	3
IA4*i*	death, illness sibling	3
IB9	fear, worry	3
IB10	rumination	3
IB12	illness, death central character	3

Card 16

IIA1*f*	contentment, parent	18
IB1	aspiration	12
IC5	aggression from environment	8
IIA2*c*	contentment with partner	8
IIA3*b*	congeniality with peer	7
IA1*a*	parental pressure	7
IA2*b*	nurturance to partner	7
IA1*i*	death or illness of parent	5
IB3	curiosity	5
IB10	rumination	5
IIA1*e*	fulfillment, parent	5
IB2	inadequacy	3
IIB2	tranquility	3
IIB7	ordinary activity	3

Card 17GF

IB5	suicide	32
IIB7	ordinary activity	32
IB3	curiosity	18
IC3	generalized restriction	12
IC1	economic pressure	10
IB1	aspiration	8
IC4	aggression to environment	8
IC5	aggression from environment	8
IA3*k*	unappreciated by peer	7
IA2*ss*	unrequited female partner	5
IB11	occupational concern	5
IB17	loneliness	5
IA1*a*	parental pressure	3
IA1*c*	nurturance to parent	3
IA2*s*	unrequited male partner	3
IA3*c*	nurturance to peer	3
IB4	behavior disorder	3
IIB5	occupational satisfaction	3

Card 18GF

IA1*b*	succorance from parent	20
IA1*a*	parental pressure	17
IA1*j*	death, illness of child	15
IA3*e*	aggression to peer	13
IA1*d*	aggression from parent	13
IB4	behavior disorder	13
IC5	aggression from environment	13
IA4*i*	death, illness of sibling	12
IC1	economic pressure	10
IA1*i*	death, illness of parent	8
IA4*e*	aggression to sibling	8
IA3*i*	death, illness of friend	7
IA4*c*	nurturance to sibling	7
IB9	fear, worry	7
IA1*c*	nurturance to parent	7
IA1*l*	disappointment in parent	7
IA1*g*	concern for child	5
IA3*b*	succorance from peer	5
IB7	guilt	5
IB12	physical illness, death of central character	5
IA1*f*	departure from parent	3
IA2*m*	jealousy of partner	3
IA2*oo*	unfaithful male partner	3
IA4*a*	pressure from sibling	3
IA4*g*	concern for sibling	3
IB3	curiosity	3
IB13	retribution	3
IB20	acquisition	3

Card 19

IC5	aggression from environment	70
IIA1*f*	contentment with parent	23
IB27	supernatural	20
IB9	fear, worry	8
IB4	behavior disorder	7
IB3	curiosity	5
IB17	loneliness	5
IC1	economic pressure	5
IIA3*b*	congeniality with peer	5
IIB2	tranquility	5
IA2*q*	pursuit of partner	3
IA3*b*	succorance from peer	3
IB2	inadequacy	3

Card 20

IB19	vacillation	38
IB17	loneliness	25
IC4	aggression to environment	17
IA3*b*	succorance from peer	15
IC5	aggression from environment	13
IB11	occupational concern	8
IB10	rumination	7
IA2*s*	unrequited male partner	7
IA3*e*	aggression to peer	7
IA3*k*	unappreciated by peer	7
IB3	curiosity	7
IB8	drunkenness	7
IB9	fear, worry	7

IA3*c*	nurturance to peer	5
IA3*j*	belongingness	5
IB2	inadequacy	5
IB14	reminiscence, sad	5
IC6	war	5
IIA2*c*	contentment	5
IIB7	ordinary activity	3
IC1	economic pressure	3
IB6	moral struggle	3

Table 5

Unusual Formal Characteristics Contributed by One or More Subjects for Each of 20 Cards

Code	Characteristics	Percentage frequency
Card 1		
III I	alternate themes	3
III G	autobiographical	2
III J	comments	2
Card 2		
III D	unreal	7
III I	alternate themes	3
III G	autobiographical	2
III J	comments	2
Card 3GF		
III I	alternate themes	5
III K	denial	3
III R	reification	3
III G	autobiographical	2
Card 4		
III D	unreal	10
III I	alternate themes	5
III K	denial	5
Card 5		
III I	alternate themes	3
III H	continuations	2
III K	denial	2
Card 6GF		
III I	alternate themes	5
III D	unreal	2
Card 7GF		
III I	alternate themes	5
III G	autobiographical	3
Card 8GF		
III D	unreal	3
III F	central character not in picture	3
III I	alternate themes	3
III R	reification	3
III C	descriptive	2
III H	continuations	2
Card 9		
III I	alternate themes	10
III G	autobiographical	2
III C	descriptive	2
III R	reification	2
Card 10		
III G	autobiographical	3
III J	comments	2
III R	reification	2
Card 11		
III C	descriptive	17
III D	unreal	13
III I	alternate themes	10
III E	fairy tale	8
III G	autobiographical	3
III R	reification	3
III A	symbolic	2
Card 12F		
III D	unreal	13
III I	alternate themes	8
III C	descriptive	7
III E	fairy tale	7
III K	denial	3
III G	autobiographical	2
III R	reification	2
Card 13MF		
III I	alternate themes	12
III G	autobiographical	2
III R	reification	2
Card 14		
III I	alternate themes	3
III H	continuations	2
III G	autobiographical	2
Card 15		
III D	unreal	10
III F	central character not in picture	7
III I	alternate themes	7
III C	descriptive	2
III J	comments	2
III K	denial	2
III R	reification	2
Card 16		
III G	autobiographical	22
III D	unreal	3
III C	descriptive	2
III F	central character not in picture	2
III H	continuations	2
III R	reifications	2
Card 17GF		
III K	denial	12
III I	alternate themes	7
III A	symbolic	3
III J	comments	3
III R	reification	3
III C	descriptive	2
III G	autobiographical	2
III L	rejection	2
Card 18GF		
III I	alternate themes	7
III K	denial	3
III G	autobiographical	2
III H	continuations	2

Card 19		
III D	unreal	17
III C	descriptive	12
III A	symbolic	8
III E	fairy tale	7
III J	comments	5
III F	central character not in picture	3
III G	autobiographical	3
III I	alternate themes	3
Card 20		
III I	alternate themes	8
III G	autobiographical	3
III R	reification	3
III A	symbolic	2
III K	denial	2

Table 6

Perceptual Distortions Contributed by One or More Subjects for Each of 20 Cards

Distortion	Percentage frequency
Card 1	
broken string	3
boy's eyes closed	2
Card 2	
*pregnancy noted	8
background is girl's imagination	3
Card 4MF	
woman in background seen as real, not picture	10
*woman in foreground seen as nurse	5
woman in foreground seen as Negress	3
*characters seen as brother and sister	2
Card 6GF	
*husband, wife (contemporaries)	68
*older husband	10
*father, daughter	7
*father or husband	3
*Sherlock Holmes	2
*boss	2
Card 7GF	
girl is holding a baby	8
uncertain if older woman is holding book	2
Card 8GF	
sexual misidentification	2
Card 9GF	
a reflection in the water	3
Card 10	
*elderly couple	25
*young couple	13
mother, son	7
two males	3
Card 12F	
sexual misidentification (figure in foreground)	3
Card 13MF	
*nudity of woman mentioned	8
*doctor, patient	7
father, daughter	3
Card 14	
individual sitting on window ledge	3
individual climbing in window	2
Card 15	
manacled hands	8
chairs instead of crosses	2
man in background	2
Card 17GF	
sexual misidentification	18
Negro character on bridge	5
don't mention character on bridge	2
seen as two different scenes	2
Card 18GF	
sexual misidentification	28
Card 20	
sexual misidentification	3
no head on man	2
no character seen	2

* These are not actually perceptual distortions. They are in this table in order to furnish normative data for unusual interpretations.

Table 7

Percentage of Subjects Contributing in a Protocol of 20 Stories a Given Number or More Stories of Each Emotional Tone Rating

Rating	Frequency in 20 stories														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
−2	98	90	73	33	15	5	3	2							
−1	100	100	100	97	92	88	75	60	40	22	15	5	2	2	2
0	100	93	87	80	57	42	33	25	15	8	3	2	2		
+1	100	73	52	37	17	10	7	3							
+2	28	2													

Table 8

Percentage of Subjects Contributing in a Protocol of 20 Stories a Given Number or More Stories of Each Outcome Rating

Rating	Frequency in 20 stories															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
−2	82	60	28	20	12	5	3	3	2	2	2	2				
−1	92	78	53	32	22	10	3									
0	82	67	37	17	10	8	7	3	2							
+1	97	90	83	73	62	45	38	27	15	12	7	5	5	3	3	2
+2	70	38	27	22	13	8	8	5	3							
?	40	13	7	2	2	2	2	2	2							
x	87	77	65	48	38	27	18	17	12	10	8	5	5	2		

Table 9

Percentages of Subjects Contributing in a Protocol of 20 Stories a Given Number or More of Every Theme Contributed at Least Twice by Any Subject

Code	Theme	Frequency in 20 stories					
		1	2	3	4	5	6
IA1*a*	parental pressure	87	62	48	20	5	2
IA1*b*	succorance from parent	77	13	3			
IA1*c*	nurturance to parent	25	3	2	2		
IA1*f*	departure from parent	18	3				
IA1*g*	concern for child	42	15	3			
IA1*i*	death or illness of parent	32	8	2			
IA1*j*	death or illness of child	45	8	2	2		
IA1*k*	disappointment to parent	23	5	3			
IA1*l*	disappointment in parent	32	5	2			
IA1*m*	filial obligation	25	3				
IA2*a*	pressure from partner	60	22	3			
IA2*b*	succorance from partner	42	12				
IA2*c*	nurturance to partner	45	12				
IA2*f*	departure from partner	38	8				
IA2*gg*	concern for female partner	12	2				
IA2*h*	illicit sex	35	5				
IA2*i*	illicit sex with violence	7	2				
IA2*j*	death or illness female partner	75	23				
IA2*jj*	death or illness male partner	23	2				
IA2*m*	competition for partner	30	5				
IA2*oo*	unfaithful to partner	30	3				
IA2*q*	pursuit of partner	18	2				
IA2*r*	seduction	27	3				
IA2*s*	unrequited male partner	18	2				
IA2*ss*	unrequited female partner	33	7				
IA2*t*	restraint	10	2				
IA2*u*	childbirth	10	2				
IA3*a*	pressure from peer	27	3				
IA3*b*	succorance from peer	30	5				
IA3*d*	aggression from peer	8	2				
IA3*e*	aggression to peer	32	8	2			
IA3*i*	death or illness friend	28	5				
IA3*j*	belongingness	47	10	3			
IA3*k*	unappreciated by peer	33	7	2			
IA3*n*	envy	8	2				
IA4*c*	nurturance to sibling	8	3				
IA4*f*	sibling rivalry	28	3	2			
IB1	aspiration	83	52	33	10		
IB2	inadequacy	47	13	3	2		
IB3	curiosity	80	45	15	2		
IB4	behavior disorder	62	12	7	5		
IB5	suicide	47	8	2	2		
IB6	moral struggle	17	2				
IB7	guilt-remorse	63	32	15	5	2	
IB9	fear-worry	60	37	8			
IB10	rumination	28	2				
IB11	occupational concern	42	37	12	7		
IB12	illness or death of central character	30	2	2			
IB14	reminiscence, sad	37	8				
IB15	intra-aggression	30	2				
IB16	religion	33	8	2			
IB17	loneliness	52	22	7	2		
IB19	vacillation	43	3				
IB20	acquisition	20	3	2			
IB21	exhaustion	18	2				
IB22	inconsequential sadness	17	2				
IB27	supernatural	30	3				
IC1	economic pressure	47	23	17	5	3	2
IC2	legal restriction	12	2				
IC3	general restriction	30	7				
IC4	aggression to environment	45	12				
IC5	aggression from environment	88	53	25	13	5	
IC6	war	28	8				
IC7	escape	50	15				

IIA1*a*	cooperation with parent	18	3	2		
IIA1*f*	contentment with parent	40	8	2		
IIA1*g*	ordinary familial activity	17	3			
IIA2*c*	contentment with partner	50	5	2		
IIA2*d*	reunion with partner	8	2			
IIA3*b*	congeniality with peers	20	2			
IIB1	self-esteem	13	5	2	2	
IIB2	tranquility	23	5			
IIB3	reminiscence, happy	40	10	2		
IIB7	ordinary activity	57	18	8	3	2
IIC1	favorable environment	23	3			

On only one out of 20 cards was there a difference in emotional tone rating between the two groups of subjects which was significant at or beyond the .05 level of confidence. Table 2 sets forth the card-by-card normative data for emotional tone. As regards outcome there were four significant card differences for the linear scale of values and no significant differences on any card for the (?) category. The outcome data are presented for each card in Table 3. In merely 10 instances were there significant differences in themes (total possible number of comparisons was 150×20=3000) and the card-by-card normative data appear in Table 4. There was a significant difference in unusual formal characteristics on only one card and there were no such differences on any single card for perceptual distortions. The normative data for these two variables are found in Tables 5 and 6 respectively.

When the data for all 20 cards were pooled, so that analysis was possible in terms of the frequency of individual subjects contributing a given number of a specific theme, emotional tone, distortion, formal characteristics, etc., in a total protocol of 20 stories, the two groups of subjects were still found to be quite similar. There were no differences in emotional tone of the stories, but the pregnant subjects did give significantly more happy outcomes than the college girl subjects (normative data in Tables 7 and 8 respectively). Out of a check list of 150 themes, on only one do these two groups differ—more college girls gave more themes of occupational concern than did the pregnant women (Table 9). There were no differences in unusual formal characteristics (Table 10) but more pregnant women produced more perceptual distortions, excluding sexual misidentifications, than did the college girls (Table 11).

From these data concerning significant differences it is apparent that both samples are really drawn from the same population insofar as TAT stories analyzed according to this methodology are concerned, since there were

Table 11

Percentages of Subjects Having a Given Number or More of Perceptual Distortions in a Protocol of 20 Stories

Type of distortion	Frequency in 20 stories				
	1	2	3	4	5
Sexual misidentifications	50	12	2		
Other distortions	57	32	5		
Total distortions	70	45	12	5	2

Table 10

Percentages of Subjects Contributing a Given Number or More Stories with Unusual Formal Characteristics in a Protocol of 20 Stories

Code	Characteristics	Frequency in 20 stories										
		1	2	3	4	5	6	7	8	9	10	11
IIIA	Symbolic	12	3									
IIIC	Descriptive	28	12	5	2	2						
IIID	Unreal	53	18	7	2							
IIIE	Fairy tale	22	2									
IIIF	Central character out	10	3									
IIIG	Autobiographical	22	5	3	2	2	2	2	2	2	2	
IIII	Alternate themes	45	25	17	10	2	2	2	2	2	2	2
IIIJ	Comments	10	2	2								
IIIK	Denial of theme	22	7	2								
IIIR	Reification	10	3	3	3	3	3	2				

hardly more than differences between the two groups than would be expected by chance in the number of implicit and direct comparisons made for each variable. Actually, then, it is reasonable to pool these data and offer norms based on a sample of 60 "normal" American women in this age range. By referring to these norms it will be possible to determine whether any given story contributed by a woman in response to a particular card is unusual or if there is a characteristic consistency throughout a series of 20 stories which is outside normal limits. These are the basic data—the frequency and unusualness of specific fantasy content — which must be determined before any interpretive significance can be placed on the thematic material obtained. How any TAT analyst will then use these data to interpret personality structure and content will depend on his own theoretic orientation and on the amount of behavioral data obtained from other sources which are available to him and in the light of which he can ascribe meaning to the fantasy material.

Received December 11, 1952.

References

1. Cohen, L. D. Stimulus properties of individual cards on the Thematic Apperception Test. Unpublished manuscript. Duke Univer., 1952.
2. Eron, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. consult. Psychol.*, 1948, **12**, 387–395.
3. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).
4. Eron, L. D., & Ritter, Anne M. A comparison of two methods of administration of the Thematic Apperception Test. *J. consult. Psychol.*, 1951, **15**, 55-61.
5. Eron, L. D., Terry, Dorothy, & Callahan, R. The use of rating scales for emotional tone of TAT stories. *J consult. Psychol.*, 1950, **14**, 473-478.
6. Freedman, L. Z., Redlich, F. C., Eron, L. D., & Jackson, Edith. Training for childbirth: remembrance of labor. *Psychosom. Med.*, 1952, **14**, 439-452.
7. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
8. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge: Harvard Univer. Press, 1943.
9. Ritter, Anne M., & Eron, L. D. The use of the Thematic Apperception Test to differentiate normal from abnormal groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 147-158.
10. Rosenzweig, S. Apperceptive norms for the Thematic Apperception Test: I. The problem of norms in projective methods. *J. Pers.*, 1949, **17**, 475-482.
11. Terry, Dorothy. The use of a rating scale of level of response in TAT stories. *J. abnorm. soc. Psychol.*, 1952, **47**, 507-511.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Ambiguity of Pictures and Extent of Personality Factors in Fantasy Responses[1]

Douglas T. Kenny
University of British Columbia

and Sidney W. Bijou
University of Washington

This paper is the second in a series reporting on the relationship between the ambiguity qualities of picture stimuli and extent of personality factors in the elicited fantasy. The first [2] described the method used for establishing ambiguity values for Thematic Apperception Test pictures. It is the object of this paper to describe the findings on the relationship between stimulus ambiguity of Thematic Apperception Test pictures and extent of personality factors evaluated from fantasy productions.

We have pointed out [2] that a common assumption in social and clinical psychology is that the more ambiguous the stimulus the greater the apperceptive distortion, hence the greater the amount of personality factors revealed. The problem of how the stimulus structure affects the extent of personality factors "determining" a response is central to any behavior theory attempting to interrelate stimulus ambiguity on one hand, and perception, apperception, and motivational states on the other. Most experiments in this area attempt to show how "inner" factors (intervening variables) play a role in perception and apperception rather than how configurational ambiguity is related to the extent of personality factors "injected" into the stimulus situation.

The hypothesis under consideration has provided part of the rationale for the construction of the TAT and other projective techniques. And because of the obvious pertinence of this hypothesis to clinical, social, and personality theory and to the construction of clinical instruments as well, there is a continual need for research which aims to evaluate its validity.

[1]The authors wish to express their appreciation to Dr. Robert Quinn and Dr. Ira Steisel for so conscientiously performing the arduous task of judging the protocols.

Method

Ascertaining ambiguity level. The first paper [2] in this series may be referred to for a discussion of alternative methods for measuring the ambiguity qualities of stimulus material for obtaining apperceptive responses. We will recapitulate briefly our approach to the problem. Data on the ambiguity of 21 TAT cards (the general and male series) were obtained by having 51 judges rank order the cards according to our definition of ambiguity. We defined ambiguity level in terms of the estimated number of interpretations a group of judges thought would apply to each TAT card. An ambiguity value for each card was therefore expressed in terms of its median rank-order value. The multi-judge reliability of the ranking was adequate ($\rho = .80$).

By this method we obtained three sets of five TAT pictures each, varying in three degrees of ambiguity. Cards 12BG, 1, 2, 9BM, and 17 BM had low median rank values and are therefore classified as least ambiguous. Cards 4, 7 BM, 8BM, 12M, and 13 MF were given intermediate rank values and make up the intermediate ambiguity group. And cards 5, 15, 18 BM, 11, and 19, which had the high rank values, comprise the most ambiguous set.

Obtaining fantasy responses. The 15 cards were administered individually to 18 male college students from introductory psychology classes at the University of Washington. The

subjects were assigned at random to three comparably trained examiners, two males and one female. Fantasy reactions to the cards were obtained according to Murray's [8] procedures with certain necessary modifications. In order to present comparable test situations, test directions were standardized. The experimental design required the subject to relate a story to each card of a set on three successive days, the cards being administered in random order between and within sessions. Stories were recorded on a dictaphone machine. Each story was then typed on a single sheet of paper and assigned a code number. From our 18 subjects we obtained 270 stories, 90 for each of the three sets of cards. Before being presented to the judges, the typewritten protocols were shuffled to eliminate the influence of fixed sequential effects.

Measuring extent of personality factors. To quantify the significance of personality factors in the TAT protocols, we used a modification of Stephenson's Q-sort technique [9]. The material consisting of the 270 stories from 18 subjects was given to the judges in two groups of 135 each. The judges, two well-qualified clinical psychologists working in different mental hygiene clinics, were asked to sort the stories in terms of the amount of personality factors revealed by placing each in one of nine categories, with the following normal frequency distribution:

Score	1	2	3	4	5	6	7	8	9
Frequency	7	11	16	20	27	20	16	11	7

Stories judged to have little personality material were given a value of 1, stories with a great deal of personality material, a value of 9.

Directions to the two judges were as follows:

You are requested to rate 135 TAT stories on the basis of the amount of personality structure and dynamics (motives and motivational structure, attitudes, defenses, conflicts, self concepts, values, goals, etc.) revealed about the subject.

Please read each story carefully and then place it in one of three piles. Place in Pile 1 all those stories you believe show few or no personality factors. Place in Pile 2 all those stories you believe show some personality features (moderately rich in personality content). Place in Pile 3 all those stories you believe show considerable or extensive personality information.

Review the stories in Pile 1 (low in personality content) and select the 7 stories you believe to have the least amount of personality information. Place these in Group 1. Review the stories in Pile 3 (high in personality content) and select the 7 stories you believe to have the greatest amount of information. Place these in Group 9. Review the stories remaining in Pile 1 and select the 11 stories you believe to have a little more information than those placed in Group 1. Place these in Group 2. Review the stories remaining in Pile 3 and select the 11 stories you believe to have considerable information but not so extensive as those placed in Group 9. Place these in Group 8. Review the stories remaining in Pile 1 and place the next 16 stories in Group 3. These would be more informative than those you placed in Group 2. Review the stories remaining in Pile 3 and place the next 16 stories in Group 7. These should be relatively rich in content but not quite so informative as those placed in Group 8.

Handle the middle group the same way. First place the 20 stories with the least amount of personality information in Group 4. Then place the 20 stories with the greatest amount of information (within this group) in Group 6. The remaining 27 stories will constitute Group 5.

Results

Differences in amount of fantasy productions. The first question likely to arise is: Is there a relationship between length of story and ambiguity level of a card? A word count was made of the 90 stories obtained at each of the three levels of ambiguity. The mean number of words composing the stories from the least ambiguous cards was 980.2, for the med-

Table 1

Differences in Story Length Related to Test Administrator

Test administrator	Mean number of words	Mean difference	*SE* of mean difference	*t*
L	1272.9			
vs.		628.6	188.76	3.33**
J	644.3			
L	1272.9			
vs.		476.7	190.78	2.50*
W	796.2			
W	796.2			
vs.		151.9	146.37	1.04
J	644.3			

** Significant at 1 per cent level.
* Significant at 5 per cent level.

Table 2

Analysis of Variance of Values from a Q-Sort of Stories Based on Personality Content

Source of variation	Sum of squares	*df*	Mean square	*F*
Examiners	1036.50	2	518.75	4.76*
Subjects in same group	1636.19	15	109.08	
Total between subjects	2672.69	17		
Raters	.00	1	.00	.....
Ambiguity level	877.17	2	438.59	12.19**
Examiners × Judges	7.39	2	3.70	.....
Examiners × Ambiguity	138.66	4	34.67	.....
Ambiguity × Judges	45.50	2	22.75	.....
Examiners × Ambiguity × Judges	62.39	4	15.60	.....
Pooled subjects × Ambiguity	2692.20	75	35.90	.....
Total within subjects	3823.31	90		
Total	6496.00	107		

** Significant at 1 per cent level.
* Significant at 5 per cent level.

ium ambiguous cards 947.7, and for the most ambiguous 785.5. None of the differences between the means was statistically significant.

Our research design produced some evidence on the question of whether the length of the stories varied as a function of the behavior (personality) of the test administrators. It will be recalled that there were three examiners, one female, two males. Word counts were made of the stories obtained by each examiner. Tests of the reliability of the differences between pairs of means indicated that one of the male examiners (L) obtained reliably longer stories than the other two examiners. These findings are presented in Table 1.

Differences in content "richness" of fantasy productions. The next question, and from the point of view of this paper, the most important, is: Are there differences in personality content in the protocols as a function of the ambiguity level of the cards? Before presenting the findings bearing on this question, a brief comment on judge reliability may be helpful.

The correlation of the scores of the two judges is .68, which is statistically significant at the 1 per cent level of confidence. While the correlation does provide a measure of agreement between the raters, it does not reflect accurately the reliability of the pooled ratings used in the analysis. Applying the Spearman-Brown prophecy formula, the correlation of .68, the estimate of the reliability of the combined judgments becomes .81.

An analysis of variance was made to determine not only the effect of ambiguity level on amount of personality material revealed but also the effect of examiner and judge differences. The results of this analysis are shown in Table 2.

Two factorial effects seem to be significant: level of ambiguity of the cards and examiner differences. However, neither the judge effects

Table 3

Differences in Content Scale Value Related to Test Administrator

Test administrator	Mean	Mean difference	*SE* of mean difference	*t*
L	5.82			
vs.		1.50	.66	2.27*
J	4.32			
L	5.82			
vs.		.91	.67	1.36
W	4.91			
W	4.91			
vs.		.59	.55	1.07
J	4.32			

* Significant at 5 per cent level.

nor any of the interactions approaches significance. To determine which examiners and which sets of cards showed differences a further analysis was made by means of *t* tests. Since the performance of the two judges did not show significant differences, their scores were pooled. Calculations of *t* tests are, therefore, in terms of the combined judgments.

Turning first to the analysis of the examiner differences we find that there was no significant difference between the two male examiners and that one of the male examiners obtained stories similar to those obtained by the female examiner. The second male examiner's subjects related stories containing higher personality values than those of the female examiner. These findings are summarized in Table 3.

Finally, examination of the personality-score differences as related to picture ambiguity, given in Table 4, shows that the cards in the medium ambiguous set were significantly higher than those in the least and most ambiguous sets. There is no statistical difference between the mean values of the least and of the most ambiguous groups of pictures.

Table 4

Differences in Content Scale Values Related to Ambiguity Level of TAT Cards

Ambiguity level	Mean scale value	Mean difference	*SE* of mean difference	*t*
Least	4.57			
vs.		1.24	.54	2.30*
Medium	5.81			
Least	4.57			
vs.		.05	.49	.10
Most	4.62			
Medium	5.81			
vs.		1.19	.39	3.05**
Most	4.62			

** Significant at the 1 per cent level.
* Significant at the 5 per cent level.

Our material yielded one other interesting bit of information. It will be recalled that Murray [9] in constructing the TAT divided the cards into two sets of ten. Cards numbered 10 and below were believed to be more structured than those numbered 11 and above. Of the 15 cards used in this study, 7 are from the first set of 10, and 8 are from the second set of 10. To determine whether there was any difference between Murray's estimate of ambiguity and extent of personality factors revealed, the data were arranged so that a mean pooled rater judgment was obtained for the 7 cards belonging to the first set of 10 and, similarly, a mean pooled rater judgment for the 8 cards belonging to the second set of 10. The mean "personality richness value" for the cards from the first 10 cards was 5.22 and that for the cards in the second 10 cards was 4.82. The *t* value equaled .93, which was not statistically significant. Hence, when the cards are arranged according to Murray's classification, we do not find his most ambiguous cards eliciting more significant personality factors than his least ambiguous cards. The possibility that the second series of pictures often elicits more meaningful material for personality analysis may be primarily a function of the change in test directions and the influence of card sequence rather than the more "dramatic" qualities of the second series of pictures.

Discussion

This study presents interesting findings not only on the problem of stimulus ambiguity and fantasy productions but also on examiner behavior and fantasy productions. One of the male examiners obtained protocols that were longer than those secured by the other two examiners. The stories given to the same examiner were also richer in personality content than those told to the female examiner. It is apparent that interpersonal stimulations between examiner and subject had differential effects despite efforts to standardize the instructions and the conduct of the examiners. For example, the examiners were required to read the test instructions, refrain from making spontaneous comments during the tests, avoid conversation before and after the administration of the tests, record the stories on the dictaphone, and the like.

Clinical psychologists have long suspected that what is obtained from projective technique is in part a function of the personality of the administrator. Recently several investigators have demonstrated this to be so, namely,

Baughman [1], Gibby, *et al.* [5], and Lord [6] working with the Rorschach technique, and Milam [7] with the Thematic Apperception Test. The implication from such studies is obvious. Studies in this area need to be designed to control or at least to evaluate influences from this source. Unfortunately this is not a simple task. It is conceivable that such stimuli could be so subtle that they would be "blind spots" not only to the examiners but also to the investigators.

Our finding on stimulus ambiguity and extent of personality information revealed in fantasy responses when taken at face value is not consistent with the hypothesis that the two terms are related in a monotonic fashion. The five most ambiguous cards produced stories that were no richer in personality content than those from the five least ambiguous cards. Moreover, cards with median ambiguous values yielded stories with more personality information than those of the other two sets. It might be argued that this finding should not be taken as evidence refuting the monotonic hypothesis because of the way the terms were defined and measured. To this position we offer two comments. First, it is axiomatic that any obtained relationship is a function of the methods of measurements involved. If we had measured either one or both of our terms differently, we might have obtained different results.

The second point relating to this argument is the fact that the hypothesis does not include definitions of terms, operational or otherwise. Each investigator is therefore forced to define the terms in accordance with his own judgment. What is being evaluated here, then, is the hypothesis with concepts measured as described.

Let us now turn to a consideration of the findings on stimulus ambiguity and personality content. Our approach is in terms of drive strength and stimulus generalization [4]. In regard to drive strength, we may infer that the 18 male college students serving as subjects (*a*) had a medium-to-weak drive level as compared to a disturbed or deprived group; and (*b*) revealed the dominant drive as varying over a wide range. For example, the dominant drive for one individual may have been achievement, for another, sex, for still another, aggression, and so on. If this were so, we can then assume that the tendencies for perceptual responses instigated by the drives of the group were not very strong and were heterogeneous in kind.

Now let us consider the stimulus properties of the 15 cards. We could assume that the card stimuli varied in at least two dimensions: (*a*) the number of cue-constellations available for selective reactions, and (*b*) the general definiteness level of the cue-constellations. According to this formulation cards in set A would have a low frequency of cue-constellations. In addition most of these cue-patterns would be sufficiently definite to evoke similar perceptual responses from members of the group. On the other hand, the most ambiguous cards, set C, would have a high frequency of cue-constellations and most of them would vary in definiteness from cue-patterns that yield a high degree of perceptual response agreement. The cards in set B would, of course, have intermediate stimulus properties between those in set A and set C.

According to this formulation, cues in set A should produce response tendencies which compete favorably with those from the dominant drives. These stories should be "card dominated." They should consist of the symbolizing (and other) processes associated with the interpreted meanings of the cards. The hypothesis of how the story evolves is somewhat like this: In compliance with test instructions the individual selects and makes a perceptual reaction to an aspect of the card. He then relates organized verbal material pertinent to the need-satisfaction sequences, the conflicts, frustrations, gratifications, and so on. Stories based primarily on card stimuli would not be expected to contain much about the individual's dominant drive and the related personality concepts. By the same token, cards in set B should evoke perceptual response tendencies about equal in strength to those from the drives. In these situations sometimes the picture stimuli dominate, at other times the drive stimuli govern. In any case, many "perceptual errors" could be made. Numerous stories would therefore be based on the dominant drives. Such protocols should yield more information about personality than those from the pictures in set A. This is in agreement with the findings. Finally the cues in the pictures of set C should elicit weak response tendencies which would easily be over-

ridden by response tendencies of the major drives. Stories from this set of pictures should be "drive dominated" and should reveal most about personality factors. This expectation is not in agreement with the findings.

What are some of the ways of accounting for the discrepancy between hypothesis and fact? (*a*) The hypothesis needs to be modified to account for another variable. It is possible, for example, that an uncontrolled influence was operating to lower fantasy behavior in response to cards in set C, which were the most ambiguous pictures. Some investigators [3] have suggested that as situations become increasingly vague, anxiety or stress is increased. Such a situational condition may have a general inhibitory effect on the performance to the least structured cards. (*b*) The cue-patterns in the cards do not vary as assumed—increasing ambiguity is correlated with an increasing number of cue-patterns and a decreasing clarity of cue-patterns. It is conceivable that there was a low relationship between either one or both of these dimensions and the judges' estimates of number of possible interpretations [2]. (*c*) The cue-patterns do vary as assumed, but there were qualitative differences among them. For instance, cards in set C may have had a greater number of cue-patterns than those in set B, but the latter had more relating to drives and situations of an interpersonal nature. In line with this comment, it should be pointed out that there were more human figures in set B than in set C. (*d*) Finally, the judges, in accordance with their training and experience as clinical psychologists, were inclined to give greater weight to protocol contents having to do with drives and dynamics associated with behavior pathology such as sex and aggression, and to underevaluate those pertaining to the milder and more "normal" processes such as affiliation and achievement. Presumably material concerning either set of terms would be revealing about personality, but because of the judges' orientation the pathological could have rated higher than the normal. Such a bias could have had a differential effect considering the content of the cards from set to set.

Summary

This paper describes an experiment testing the assumption of a direct relationship between stimulus ambiguity and extent of personality factors revealed in evoked fantasy. On the basis of findings in an earlier study, three sets of TAT cards varying in degree of ambiguity were selected and given to 18 young adult male subjects. Two well-qualified clinical psychologists, utilizing a Q-sort technique, judged the significance of the personality factors injected into the fantasy stories.

The results from an analysis of variance of the Q-sort scores are not consistent with the hypothesis under investigation. With increasing stimulus ambiguity there is an initial increase in the extent of personality factors revealed in the fantasy, and then a decrease. In addition, it was shown that the behavior of the examiner plays a role in fantasy productions. It was also evident that with instructions held constant, cards from Murray's second series did not reveal more information about personality than cards from his first series.

The hypothesis and the findings are analyzed in terms of drive strength and stimulus generalization.

Received December 1, 1952.

References

1. Baughman, E. E. Rorschach scores as a function of examiner difference. *J. proj. Tech.*, 1951, 15, 243–249.
2. Bijou, S. W., & Kenny, D. T. The ambiguity values of TAT cards. *J. consult. Psychol.*, 1951, 15, 203–209.
3. Cameron, N., & Magaret, Ann. *Behavior pathology.* Boston: Houghton Mifflin, 1951.
4. Dollard, J., & Miller, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
5. Gibby, R. G., Miller, D. R., & Walker, E. L. Examiner variance in the Rorschach protocols of neuropsychiatric patients. *Amer. Psychologist*, 1952, 7, 337–338. (Abstract)
6. Lord, Edith. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, 64, No. 10 (Whole No. 316).
7. Milam, J. R. Examiner influences on thematic apperception stories. Unpublished master's dissertation, Univer. of Washington, 1952.
8. Murray, H. A. *Manual for the Thematic Apperception Test.* Cambridge: Harvard Univer. Printing Office, 1943.
9. Stephenson, W. Methodological considerations of Jung's typology. *J. ment. Sci.*, 1939, 85, 185–205.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Measuring Achievement Motivation with "Picture Interpretations"

Henry H. Morgan
Psychological Corporation

The measurement of achievement motivation is the subject of recent research which has been reported by McClelland and others [2, 3, 4, 6]. This research employed a short version of the Thematic Apperception Test (TAT) in which stories written in response to picture slides were scored for Murray's need for Achievement, or *n Ach* [7, pp. 144, 164]. In the initial studies by McClelland *et al.* [4], groups of subjects took the test under experimentally induced motivational states, and a scoring system was evolved which made use of differences found between those taking the test under "ego-involving" conditions and those taking it under "neutral" conditions.

The study about to be described is an outgrowth of the research mentioned above and is part of an Office of Naval Research project undertaken to develop a standardization test of achievement motivation.[1] One purpose of the study was to determine the relationship between scores for achievement motivation, or *n Ach*, and a criterion of achievement which, in this case, was school grades. Another purpose was to test the stability of *n Ach* scores over a period of time.

Procedure

The sample. The subjects tested were boys in the junior class of high schools in a New England industrial city. Five groups of boys were used, ranging in number from 39 to 62 subjects. Groups 1, 2, and 3 were from a large academic high school, while Groups 4 and 5 were from a vocational high school where shop courses contributed to the grade criterion.

Students had taken the Otis Quick-Scoring Mental Ability Test (Beta test, Form Cm) during their freshman year in high school, and Otis IQ's were available for a majority of the subjects. Each academic high school group had an average Otis IQ of 109, while each vocational high school group had an average IQ of 99 on the Otis.

The measuring instrument. The achievement motivation test under development has been designated "Picture Interpretations," or P-I, to distinguish it from Murray's TAT. From a pool of pictures, some of which had been pretested in other high school populations,[2] four forms of P-I were constructed. These forms, incorporating 36 different pictures, were designated E, F, G, and H. Forms E, F, and G, which were made up of 12 pictures each, used plain unstructured response sheets (8½ in. × 14 in.) for the recording of the stories by the subjects. Form H, which contained the first eight pictures of Form G, had the same over-all time limit of 30 minutes for the test but used slightly different response sheets on which were printed a series of "story structuring" questions used in previous *n Ach* studies [4]. In all groups, subjects viewed each slide projection for a period of ten seconds. In the administration of Forms E, F, and G two and one-half minutes were allowed after each slide presentation for story writing, but in the administration of Form H three and three-quarters minutes were allowed for the writing of each story.

Subjects were told that the testing program was sponsored by the Office of Naval Research, that the test results would be used for

[1]Research Project NR 172-116, Contract No. N onr 559(03), 1951-52, was under the direction of D. C. McClelland, Wesleyan University. Contract personnel included R. A. Clark, H. H. Morgan, R. W. Freeman, and R. W. Lull.

[2]Forms A, B, C, and D of P-I yielded correlations of *n Ach* scores and grades that ranged from .22 to .38. With intelligence partialed out, correlations of *n Ach* scores and grades ranged from .13 to .38.

research purposes only, and that students' responses would be kept anonymous. Each response booklet contained mimeographed instructions for the test, and these instructions were read aloud to the subjects before each test administration. The instructions for Forms E, F, and G are given below. Those for Form H were slightly different owing to the altered time limits.

Instructions

You are going to see a series of pictures, and your task is to tell a story that is suggested to you by each picture. Try to imagine what is going on in each picture. Then tell what the situation is, what led up to the situation, what the people are thinking and feeling, and what they will do.

In other words, write as complete a story as you can—a story with a plot and characters.

You will have ten seconds to look at a picture and then two and one-half minutes to write your story about it. Write your first impressions and work rapidly. I will keep time and tell you when it is time to finish your story and to get ready for the next picture.

There are no right or wrong stories or kinds of stories, so you may feel free to write whatever story is suggested to you when you look at a picture. Spelling, punctuation, and grammar are not important. What is important is to write out as fully and as quickly as possible the story that comes into your mind as you imagine what is going on in each picture.

Notice that there is a half page for writing each story and that a picture number appears on each half page. If you need more space for writing any story, use the reverse side of the paper.

In the administration of all forms a warning, "Try to finish your story in 30 seconds," was given one-half minute before expiration of the allotted writing time.

Data collection and analysis. Forms E, F, and G of P-I were administered to Groups 1, 2, and 3, respectively, in March, 1952, and each group was given either Form F or Form E as an alternate form five weeks later in April. Groups 4 and 5 were tested with Forms G and H, respectively, in February, 1952.

The protocols were analyzed for *n Ach* by two scorers using P-I scoring system "D",[3] a modification of system "C" [4]. A reliability coefficient of .89 was obtained for a sample of protocols analyzed by each of the scorers.[4]

Results

Product moment correlation coefficients between *n Ach* scores, grade averages, and intelligence test scores are given in Tables 1 through 5. Tables 1, 2, and 3 present the results for the academic high school groups, while the findings for the vocational high school groups are given in Tables 4 and 5.

Correlations in Tables 1 through 5 which are .39 or greater are significant beyond the .01 level of confidence using a CR technique [5, p. 122]. Correlations of .29 to .36 are significant beyond the .05 level. Correlations that are .27 or lower do not reach the .05 level of confidence. The partial correlations, which are between *n Ach* scores and school grades with Otis IQ partialed out [5, p. 141], give an estimate of the relationship between *n Ach* and the achievement criterion if intelligence is held constant. The subscripts of the partial *r*'s refer to the variables listed in the correspondingly numbered columns or rows in any given table.

Examination of Tables 1 through 5 reveals that, in general, low to moderate relationships

Table 1

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in an Academic High School, Sample No. 1 ($N = 53$ to 62)

(Form F administered 5 weeks after Form E)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form E)	5 *n Ach* (Form F)
1. Otis IQ	—	.42	.45	.03	.09
2. Sophomore Grade Average	.42	—	.81	.16	.19
3. Junior Grade Average	.45	.81	—	.25	.29
4. *n Ach* (Form E)	.03	.16	.25	—	.56
5. *n Ach* (Form F)	.09	.19	.29	.56	—

Partial Correlations: $r_{42.1} = .15$; $r_{43.1} = .26$; $r_{52.1} = .17$; $r_{53.1} = .28$.

[3] McClelland, D. C., Clark, R. A., and Morgan, H. H., Revisions in the system of scoring imaginative stories for need: Achievement (*n Ach*); scoring system D (April, 1952). Unpublished paper.

[4] Protocols for Groups 1, 2, and 3 were scored by the author, those for Groups 4 and 5 by R. W. Freeman. Interscorer reliability on 20 protocols from Group 3, Form G, was .89.

Table 2

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in an Academic High School, Sample No. 2 ($N = 39$ to 53)

(Form E administered 5 weeks after Form F)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form F)	5 *n Ach* (Form E)
1. Otis IQ	—	.16	.56	.24	.73
2. Sophomore Grade Average	.16	—	.65	.45	.43
3. Junior Grade Average	.56	.65	—	.47	.29
4. *n Ach* (Form F)	.24	.45	.47	—	.56
5. *n Ach* (Form E)	.73	.43	.29	.56	—

Partial Correlations: $r_{42.1} = .43$; $r_{43.1} = .42$; $r_{52.1} = .46$; $r_{53.1} = -.21$.

Table 3

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in an Academic High School, Sample No. 3 ($N = 51$ to 61)

(Form E administered 5 weeks after Form G)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form G)	5 *n Ach* (Form E)
1. Otis IQ	—	.35	.50	.22	.13
2. Sophomore Grade Average	.35	—	.70	.41	.28
3. Junior Grade Average	.50	.70	—	.39	.24
4. *n Ach* (Form G)	.22	.41	.39	—	.64
5. *n Ach* (Form E)	.13	.28	.24	.64	—

Partial Correlations: $r_{42.1} = .36$; $r_{43.1} = .33$; $r_{52.1} = .25$; $r_{53.1} = .20$.

exist between *n Ach* scores and grades. There are also low, moderate, and high relationships between *n Ach* scores and intelligence test scores as well as substantial relationships between intelligence test scores and school grades. With intelligence partialed out, relationships of *n Ach* scores and school grades tend to be somewhat reduced.

Table 4

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in a Vocational High School, Sample No. 4 ($N = 45$ to 50)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form G)
1. Otis IQ	—	.27	.32	.26
2. Sophomore Grade Average	.27	—	.70	.04
3. Junior Grade Average	.32	.70	—	.15
4. *n Ach* (Form G)	.26	.04	.15	—

Partial Correlations: $r_{42.1} = -.03$; $r_{43.1} = .07$.

Table 5

Correlations between IQ, School Grades, and *n Ach* Scores among Junior Boys in a Vocational High School, Sample No. 5 ($N = 40$ to 48)

Variable	1 Otis IQ	2 Sophomore Grade Average	3 Junior Grade Average	4 *n Ach* (Form H)
1. Otis IQ	—	.55	.46	.36
2. Sophomore Grade Average	.55	—	.67	.31
3. Junior Grade Average	.46	.67	—	.27
4. *n Ach* (Form H)	.36	.31	.27	—

Partial Correlations: $r_{42.1} = .14$; $r_{43.1} = .09$.

Tables 1, 2, and 3 give the correlations between alternate forms of P-I administered five weeks apart. These correlations of .56, .56, and .64 indicate that *n Ach* scores are fairly stable over a short period of time.

Discussion

The correlations between *n Ach* scores and grades vary from one group to another, and these variations may be due to several causes. In the first place, sample sizes are limited, ranging from 39 to 61 subjects, and therefore random fluctuations of considerable size may be expected among the coefficients computed in the various groups or samples. Some fluctua-

tions fall well within *SEr* which, in most cases, approximates .14. In addition to unreliability due to the limited size of the samples, a certain amount of unreliability is probably inherent in the grade criterion, whose reliability can be estimated by the correlations between sophomore and junior grades which range from .65 to .81. Furthermore, the reliability coefficients of .56, .56, and 64 based on alternate forms of the achievement motivation test show that the P-I technique itself is another source of unreliability. One more factor accounting for some of the variation in corrrelations from one sample to another may be the differential effectiveness of the various test forms used in measuring *n Ach*.[5]

In this study, the *n Ach* scores derived from alternate P-I forms appear, for projective tests, to be moderately reliable. However, the relationships of *n Ach* scores to grades and to intelligence appear less stable, and predictions on the basis of individual *n Ach* scores may not be made with confidence at the present stage of test development. Since achievement motivation, as measured by *n Ach* score, might be expected to vary somewhat from one time to another in the same individual, future P-I test reliabilities may not rise very much above those reported in this study.[6] These reliabilities demonstrate that *n Ach* score may be considered a measurement of an enduring personality trait just as the initial experimental studies of McClelland *et al.* [4] showed that *n Ach* score reflects temporarily induced motivational states.

A number of writers [1, 8] believe that, in work with projective techniques, standard test-retest or split-half reliabilities are of less importance than interscorer reliabilities which test the stability of interpretation from one investigator to another. The interscorer reliability of .89 reported in this study, as well as higher figures reported by McClelland [4], show that the *n Ach* scoring system is a highly consistent one.

The results of this study and other P-I studies show that the test, which began as a projective technique for the experimental study of motivation, holds some promise as a predictive instrument for forecasting achievement.

Summary

A story-telling test known as Picture Interpretations, or P-I, was administered to five groups of junior boys in academic and vocational high schools. Stories written in response to a series of pictures were scored for achievement motivation or need for Achievement (*n Ach*).

Low to moderate relationships were found between *n Ach* scores and school grades and also between *n Ach* scores and intelligence. The relationship between *n Ach* scores and school grades tended to be somewhat less when intelligence was held constant by use of the partial correlation technique.

Alternate forms of P-I given five weeks apart yielded reliability coefficients of .56, .56, and .64 indicating *n Ach* scores to be fairly stable over brief periods of time.

Received November 10, 1952.

References

1. Harrison, R., & Rotter, J. B. A note on the reliability of the TAT. *J. abnorm. soc. Psychol.*, 1945, **40**, 97–99.
2. Lowell, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1952, **33**, 31–40.
3. McClelland, D. C. *Personality*. New York: William Sloan, 1951.
4. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953, in press.
5. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
6. Morgan, H. H. A psychometric comparison of achieving and nonachieving college students of high ability. *J. consult. Psychol.*, 1952, **16**, 292–298.
7. Murray, H.A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
8. Rosenzweig, S. Idiodynamics in personality theory with special reference to projective methods. *Psychol. Rev.*, 1951, **58**, 213–223.

[5]Further analysis of the data has been undertaken by Dr. Henry N. Ricciuti, Educational Testing Service, to determine which pictures and which components of the *n Ach* scoring system are contributing consistently to the relationship between *n Ach* scores and achievement criteria.

[6]Split-half reliability coefficients (Spearman-Brown) on Forms A, B, C, and D ranged from .36 to .77.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Social and General Intelligence and Performance on the Rosenzweig Picture-Frustration Study

Lawrence Karlin
New York University

and Milton M. Schwartz
Rutgers University

Inspection of the various social situations depicted by the Rosenzweig Picture-Frustration (P-F) test suggested the possibility that social intelligence, aside from other determinants such as motivation, needs, etc., may play a role in determining the type of response that is made to these situations. The purpose of the present study is to explore this possibility by analyzing the pattern of interrelationships obtained between the scoring categories of the Rosenzweig P-F test and the subtests of a well-known social intelligence test, in this case the George Washington University Series Social Intelligence Test.

To some extent at least, responses to the various situations on the P-F test might require a certain degree of inventiveness and social insight and thus one might expect a positive relationship between the occurrence of responses indicating these characteristics on the P-F test and the subtest purportedly measuring the same characteristics on the Social Intelligence test. Furthermore, indications of behavioral difficulties obtained on the P-F test might correlate with lack of social skills as measured by the Social Intelligence test. Thus, the comparison of the pattern of performance obtained on each might enhance the understanding of the projective material furnished by the P-F test.

To complement this comparison it was decided to have a conventional test of general intelligence, the Otis Self-Administering Tests of Mental Ability. It was felt in such an exploratory study that the general hypothesis centering around the influence of social skills which might make use of a person's inventiveness and insight would also be reflected in significant relationship with a test of general intelligence.

Subjects and Materials

The three tests were administered to 45 male students in introductory psychology classes at Rutgers University at two-week intervals in the following order: Otis Test of Mental Ability, Rosenzweig P-F test, and the Social Intelligence test. The following is a brief description of the tests employed:

1. Rosenzweig Picture-Frustration Study (Revised Form) [3]. This is a semiprojective personality test consisting of 24 picture situations portraying in cartoon-like fashion typically encountered "frustrating" experiences. In each picture there is depicted conversation between the person producing the frustration and the one presumably being frustrated. The subject is required to write, in the space provided, his first response to the frustrating statement expressed. Scoring is in terms of the standard procedure of Rosenzweig which consists of seven scoring categories; three involve Direction of Aggression (Outward, Inward, and Nonaggression) and four Types of Reaction, (Obstacle-dominance, Ego-defense, Need-persistence, and Denial).

2. The George Washington University Series Social Intelligence Test of Moss, *et al.* [1]. This test consists of five subtests, namely, Judgment in Social Situations, Recognition of the Mental State of the Speaker, Memory for Names and Faces, Observation of Human Behavior, Sense of Humor. Each test may be scored separately and the scores may be combined into a total score which represents the sum of the subtests. These will be described under results wherever necessary to elucidate the meaning of the findings of this study. Suffice it to say at this point that the literal meanings of each of the five titles provides a good idea of the content of each.

3. The Otis Self-Administering Tests of Mental

Table 1
Chi Squares for the P-F, Social, and Otis Tests

Variable	*df*	P-F Test Categories							Otis test
		Extrapunitive	Intropunitive	Impunitive	Obstacle-dominance	Ego-defense	Need-persistence	Denial	
Social intelligence									
Judgment in social situations	1	.47	−.08	2.00	.16	.16	−2.33	.04	.02
Recognition of mental state of speaker	2	−11.30***	4.85*	7.48**	2.02	−.94	.76	−.67	3.10
Memory for names, faces	2	−3.38	6.59**	−.20	−1.80	−.29	−1.23	20.80***	.01
Observation of human behavior	1	.63	−1.40	−.41	.00	−.29	1.11	−.41	.04
Sense of humor	2	.40	−1.54	−.17	.09	1.14	−1.65	.09	.90
Total	1	−.99	.23	.22	.08	−.15	.00	−.09	.08
Otis test	1	.36	.36	−.36	11.00***	−.36	22.60***	−.70	

Note—Negative sign indicates that the trend of the relationship is inverse.
* Significant between .05 and .10 level.
** Significant between .01 and .05 level.
*** Significant beyond .01 level.

Ability, Higher Examination, Form D [2]. This well-known test consists of 75 heterogeneous items relating to such aspects of general intelligence as measured by number series, analogies, reasoning, etc. The shorter 20-minute time limit was used with this test.

Results and Discussion

The major results are summarized in Table 1 which shows the interrelationships among the scoring categories of the Rosenzweig P-F test, the subtests of the Social Intelligence test, and the Otis Intelligence test. The entries on this table are chi squares determined from contingency tables in which the scoring categories of the P-F test were divided into two categories by median split while some of the subtest scores of the Social Intelligence test were divided by median split and others into high, middle, and low categories. The latter type of division was necessary because in some of these subtests too many scores were concentrated at the median. Thus some of the responses are based on 2×2 and some on 3×2 contingency tables.

Inspection of these tables discloses that most of the values are not significant. However, there are more highly significant chi squares than would be expected in a random sample of such chi squares.

The salient feature of this table lies in the fact that the major scoring categories of the P-F test (Extrapunitive, Intropunitive, Impunitive) yield two highly significant chi squares and one near-significant chi square with the subtest entitled Recognition of the Mental State of the Speaker (RMSS) on the Social Intelligence test. To evaluate these relationships let us first describe briefly the types of items in this subtest and in the P-F test. The subtest RMSS consists of eighteen 4-choice multiple-choice items in which the alternative that best describes the mood of a given quotation has to be selected. One item, Number 16, is as follows:

"If only I had not let that opportunity slip I might now be enjoying the things I had longed for."

(A) Despair (B) Disappointment (C) Loneliness (D) Regret

A comparison of this type of item with the items in the P-F test suggests a basic similarity in that the P-F test item presents a situation which provides information that may be used

to characterize the mental state of one of the speakers. For example, in Item 2 there is shown a situation in which fragments of a vase lay scattered on the floor near a table and a woman is seen saying to another woman "How awful! That was my mother's favorite vase you broke." Obviously the testee provides the response for the other person. It is clear that the statement as made in the context depicted by the cartoon provides a basis for assessing the mental state of the speaker. It should be noted that this type of item provides the information about the mental state of the speaker on the basis of *both* verbal and pictorial material as against only the verbal material provided by the Social Intelligence test.

The question with which we are specifically concerned is whether the ability to recognize the mental state of the speaker as measured by the Social Intelligence test is importantly involved in the response made to the P-F test type of item. The aforementioned significant and near-significant chi-square values between the main scoring categories of the P-F test and this subtest suggest an affirmative answer to this question. This conclusion is further documented by the fact that a fairly clear and specific rationale can be found for these relationships.

Thus, the Extrapunitive (*E*) category yields a chi-square value significant at better than the .01 level. In addition, the relationship is inverse in the sense that a high *E* score tends to go with a low RMSS score. The meaning of this result seems to be fairly straightforward. The high-*E* individual is typically very outwardly aggressive and tends not to modify his responses in accordance with the needs of others. If he did, it would necessarily limit the outward expression of his aggression. It should be cautioned at this point that there are other dimensions (*N-P, O-D,* etc.) in the description of the *E*-type person. We have here emphasized the most important and pertinent dimension for the purpose of this comparison.

The second most significant chi square obtained is between the RMSS category of the Social Intelligence test and the P-F test Impunitive (*M*) scoring category. The high-*M* individual tends to be fairly objective in his reactions to situations in that he does not react in an aggressive manner either outwardly or inwardly; he does not, in general, place blame for the difficulty and tries to minimize the frustrating features of the situation. This type of behavior may be due to a lack of concern for the situation (in extreme cases, a type of schizoid reaction), or it may be due to a desire to protect the feelings of all involved. It is this last characteristic, we feel, that provides the basis for the positive relationship between the RMSS category and this one on the assumption that the desire to protect the feelings of others arises from an increased awareness of and sensitivity to their mental states.

Finally, the Intropunitive (*I*) category of the P-F test displays a tendency to vary positively with the RMSS score. Chi square is significant here between the .05 and .10 levels so that by conventional statistical standards the validity of this relationship is more speculative than were the other two. However, it is not difficult to find a rationale for this positive association. The high-*I* person is often one who tends first to blame himself when involved with frustrating problems. We may hypothesize that a basis for this self-blame lies partly in a developed appreciation and exaggerated sensitivity to other people's reactions.

The other Social Intelligence test scoring category to show significant relationship to the P-F test is Memory for Names and Faces (MNF). At the outset of the test the subjects are required to study for four minutes the photographs of the faces of twelve college-age males with a name below each face. Midway in the test the subjects are shown twenty-five faces among which are distributed the twelve originally examined. For each of the twelve names four of the twenty-five faces are suggested from which the subject must choose the correct face associated with the given name. MNF is significantly related to the Intropunitive (*I*) and Denial (*D*) scoring categories of the P-F test.

Examination of the two types of "items" subsumed under the categories of the respective tests reveals no palpable resemblances. In fact the facial characteristics of the cartoon figures in the P-F test are deliberately left unstructured to enhance its projective features (sex and age, however, are discernible).

Thus, unlike the comparisons discussed above, the form and content of these two types of items do not provide a ready rationale for

explaining the high positive association indicated by the very significant chi squares obtained. On the basis of the conventional descriptions as well as our own experience, a rather speculative hypothesis to account for this association is offered. Specifically the Denial (*D*) category is symptomatic of generalized ego-insecurity. In fact, the high *D* is specifically suggestive of this characteristic. Our experience with people who are predominantly characterized by ego-insecurity suggests that these tend to be alerted to the facial expressions of others to a high degree. The reason for this is primarily due to their need for a continual assessment of their position in a social situation and this they must do by observing the behavior of others. The *D*-type person, being very anxious, tends to rehearse the events of previous situations including the observed behavior of people in these situations. This is a symptom which often goes with marked anxiety such as is created by insecurity. It is hypothesized, therefore, that the excessive attention to and rehearsal of others may have developed his ability to recognize faces. Thus, in the MNF situation at least a part of the task, that is, recognizing a face, is facilitated. Admittedly the empirical links between this hypothesis and performance on the two types of items are rather indirect. Still, it would not be impossible to test this hypothesis by correlating the MNF subtest with some other measure of feelings of inadequacy.

It may be noted also that the Intropunitive (*I*) category shows significant relationship with MNF ($<.05$). The writer can offer no basis for this other than to suggest the possibility that the *I* and *D* categories may be correlated.

Turning now to the interrelationships between the Otis Test of Mental Ability and the P-F test, we find two striking results, namely, highly significant chi squares for the Obstacle-Dominance (*O-D*) and Need-Persistence (*N-P*) categories. High score in the *O-D* category indicates, other things equal, an individual who emphasizes the frustrating character of the situation, that is, the particular circumstance which is at the root of the frustration. Thus, he is highly task-oriented. It may be that the significant chi square in this instance indicates that the more intelligent subject tends habitually to emphasize the problem aspect of the situation rather than the personalities involved in the situation. In short, it may be that the more intelligent individual, because of his greater success with problems, exhibits a tendency to be task-oriented rather than ego-oriented. A similar rationale may be employed in the case of the *N-P* category. In fact, to an even greater extent this type of individual is concerned with solving the problem involved in the P-F test.

Summary and Conclusion

The present study compared performance on three tests, the Rosenzweig P-F, George Washington University Social Intelligence, and the Otis Mental Ability tests. The purpose of this comparison was, in part, to determine the relationship between the influence of such abilities as are measured in the two types of intelligence tests on performance in the P-F test. The significant relationships which were found suggest the possibility that certain ability factors are related to the type of response made on the P-F test. This finding also suggests the general hypothesis that differences in the way the P-F test situations are perceived partly condition responses made to the test.

Received November 10, 1952.

References

1. Moss, F. A. The measurement of social intelligence. *Psychol. Bull.*, 1928, **25**, 168–169.
2. Otis, A. S. *Otis self-administering tests of mental ability* (manual). Yonkers, N. Y.: World Book Co., 1928.
3. Rosenzweig, S. The picture-association method and its application in a study of reactions to frustration. *J. Pers.*, 1945, **14**, 3–23.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

The Practical Utility of the Clinical Report

Charles A. Dailey
Veterans Administration[1]

This paper will present a technique of measuring the practical utility of a clinical report. It is an attempt to find an objective way of answering the question, "How should patients or clients be described in the psychological report?"

A recent empirical study by Davenport [3] dealt with this problem from the standpoint of the relationship of the clinical report to the data upon which it is based. In contrast, the present technique is concerned with the relationship of the clinical report to the treatment process itself.

For purposes of the present discussion, the *utility* of a report is defined as relevance to possible decisions about treatment of the patient described in the report. In measuring utility, one must compare the report with the treatment processes employed in the institution in which the report is to be used.

Since a report is always made *to* someone, semantic *clarity* is also involved in appraising its practical value. It is essential to study clarity for the reason that it is possible to say something relevant about a patient without saying it clearly.

The Measurement of Utility

The present method of measuring utility requires the following steps.

1. We first obtain a list of the decisions which are ordinarily made by the particular counseling, neuropsychiatric, or psychological service. In the exploratory study described here, we requested the psychologists, psychiatrists, and social workers of our service to recommend decisions about treatment which, in their opinion, are frequently made concerning the patients of their service. After refining these recommendations, we obtained the list shown in Table 1. It will be noted that this list contains a number of decisions which probably have no relevance to psychological reports. They are nevertheless included for the purpose of illustrating the technique. It is also obvious that the list is only a crude approximation to the decision-making process as it actually operates on such a service.

2. As a second step, the list was submitted to a group of ten clinical psychologists familiar with this service. Each was asked to "make the set of 32 decisions as if you were dealing with the *average* neuropsychiatric patient on this service. Indicate your opinion as to the appropriate treatment for this hypothetical average patient by stating 'yes,' 'no,' or 'don't know' in response to each question." The responses of these clinicians were compiled and are reported in Table 2. This consensus of clinical opinion provides a base line from which the new information value of a report may be calculated. Such a measure is important to provide a control upon stereotypy in decision making; we must make certain that, in making decisions about a patient after reading a report about him, the clinician is not merely repeating a priori decisions.[2]

3. The third step requires the appraisal of the reports which are to be measured for utility. A clinician who is otherwise unfamiliar with the specific patient reads each report and then fills out, in checklist form, his recommendations for each of the decisions about treatment. His recommendations are then scored as to their deviation from the a priori recommendations (Table 2). A weighting system is employed for this purpose which attributes to each decision the inverse of the proportion of clini-

[1]From the Veterans Administration Hospital, Jefferson Barracks, Missouri.

[2]The author is indebted to the work of Cronbach [1] and of Kelly and Fiske [5] for emphasis upon the concept of relative information value in psychological appraisal.

Table 1

List of Frequently Made Clinical Decisions

1. Should the patient have been admitted to the psychiatric ward?
2. Should he be put on closed ward (for reasons other than electroconvulsive therapy)?
3. Should he receive electroconvulsive therapy?
4. Should he receive subshock insulin?
5. Should he receive intensive, interpretive individual psychotherapy?
6. Should he receive supportive, noninterpretive psychotherapy?
7. Should he receive group psychotherapy?
8. Should he be restricted for the first three days?
9. Should he receive repressive individual psychotherapy (i.e., therapy directed toward inhibiting completely some impulses and ideas)?
10. Should he be allowed visitors now from immediate family?
11. Should he receive sedative drugs?
12. Should he receive vitamins?
13. Should he be placed on a special diet?
14. Should he receive placebo?
15. Should he have amytal interviews for cathartic purposes?
16. Should he have hypnotic treatment for cathartic purposes?
17. Should he have a male therapist?
18. Should the therapist play an authoritarian role?
19. Should the therapist play a nurturant role (i.e., kindly and especially tolerant of dependency needs)?
20. Should the patient have occupational therapy?
21. Should this patient discuss his vocational future with the vocational counselor?
22. Should the patient work in Manual Arts Therapy?
23. Should the patient be transferred now to a mental hygiene clinic?
24. Should his relatives be given therapeutic interviews?
25. Should the patient receive especially solicitous attention from ward personnel?
26. Should the patient be particularly encouraged about his resources for "adequate living," "success," etc.?
27. Should the patient be watched for assaultive tendencies?
28. Should the patient be watched for suicidal tendencies?
29. Should the patient be encouraged to make his own decisions about length of stay?
30. Should the patient be encouraged to improve his educational status?
31. Should the patient be encouraged to change his job?
32. Should the patient be encouraged to seek the aid of Alcoholics Anonymous?

Table 2

Responses of Ten Clinicians to the Question of What Treatment Should be Given to the Average Neuropsychiatric Patient

Treatment decision number	Number of clinicians responding		
	Yes	?	No
1	9	1	0
2	0	4	6
3	0	5	5
4	0	6	4
5	2	3	5
6	8	2	0
7	9	0	1
8	9	1	0
9	1	4	5
10	7	2	1
11	2	6	2
12	6	3	1
13	1	5	4
14	2	5	3
15	0	3	7
16	0	2	8
17	6	4	0
18	0	5	5
19	4	5	1
20	10	0	0
21	6	2	2
22	8	2	0
23	1	6	3
24	7	2	1
25	6	1	3
26	6	1	3
27	4	2	4
28	3	2	5
29	7	2	1
30	0	4	6
31	0	3	7
32	0	5	5

cians recommending that treatment for the average patient. To derive this system, each number in Table 2 is subtracted from ten; the remainders are a set of weights. The scores are summed for the 32 decisions; the resulting total score represents the tendency of the report to suggest to the clinician decisions which would not have been made on the basis of "zero" information about the patient.

4. If the total score is divided by ten (no transformation is necessary if proportions are used in Step 3, instead of raw number of clinicians, to derive the weights), the resulting number represents the number of the 32 decisions which were "completely novel." This is, of course, a hypothetical number which assumes

that two "half-novel" decisions should be equated to one "completely novel" decision (i. e., one which is made exactly opposite to the a priori expectation). This variable allows us to compare various weighting systems and to make full use of our experimental data.

We shall hereinafter refer to this measure as *number of new, useful decisions.* Table 3 presents a distribution of such measures, in which the decisions of two judges for each patient are used. Since two judges are used, this modification of the measure is called "number of *clear*, new, useful decisions," for reasons which will be elaborated in the following sections.

The Measurement of Clarity

The clarity of a report is defined as the amount of agreement of two independent readers. As we are concerned with utility, it is understood that the agreement is to be about *treatment* (as distinguished from, for example, the more esoteric categories used in studying reliability of personality judgments). Clarity is appraised in the following steps.

1. Two clinicians read a given report independently. The report constitutes their sole information about the patient.

2. Following the reading, they are requested to fill out the decision check list (Table 1) independently, responding "Yes," "No," or "Don't know" with regard to each decision.

3. Their decisions are compared and a score, *number of clear, useful decisions* (the number upon which they agree), is calculated. A distribution of such scores is presented in Table 3.

The measurement of clear, new, useful decisions. It is desirable to compute an over-all value which combines the opinions of separate judges, and which also describes the departure of their opinions from a priori convictions about treatment.

For this purpose, the weighting system described above (under *The Measurement of Utility*) is applied to those decisions upon which two clinicians separately concur. The resulting measure is termed *number of clear, new, useful decisions.*

The Clarity and Utility of Nine Psychological Reports

Nine clinical reports were selected at random from our files. The reports were distributed to seven clinicians familiar with our treatment procedures. They were distributed in such a way that each report was read by two clinicians and so that no two clinicians read more than one report in common.

Table 3
Clinical Utility and Clarity in Nine Reports

Patient	Number of clear, useful decisions	Number of new, clear, useful decisions	Proportion of clear decisions that are new
A	19	9.4	.50
B	18	10.8	.60
C	15	7.8	.52
D	22	12.0	.54
E	14	8.3	.60
F	11	5.2	.47
G	11	5.1	.46
H	22	9.5	.43
I	17	7.2	.42
Median	17	8.3	.50
Per Cent of Total*	.53	.26	

* The total is 32 decisions.

On concluding his study of the report, each clinician was requested to make the 32 decisions (Table 1) for the patient described in the report. The responses were analyzed according to the operations described above. Results are summarized in Table 3. In interpreting Table 3, we can state that slightly over half of the 32 decisions can be clearly agreed on after reading the report. Of those which can be clearly agreed on, only half represent decisions which would not have been made before reading the report (according to Table 2).

Thus, about one fourth (26%) of the list of therapeutic decisions are clearly indicated contrary to a priori expectations in the average psychological report, within this sample of reports. It is, of course, not possible to say whether this is a "satisfactory" or an "unsatisfactory" proportion, with our present level of knowledge about the proper use of these treatments.

We can, however, go on to consider the

problem of which features of the psychological report may influence its clarity and utility.

Linguistic Correlates of Clarity and Utility[3]

The clinical reports which were measured above as to clarity and utility were also measured as to several factors which describe style and properties of the writing in each report:

1. length of report (in words)
2. average sentence length (in words)
3. average number of technical terms (per word)
4. average number of personalizing terms (per word)

The variables "technical terms" and "personalizing terms" require definition. The author selected from a large number of psychological reports those words which appeared to be frequently found in the vocabulary of psychologists but not in the conventional speech of lay persons (e. g., "dysfunction"), or which, while appearing in the speech of lay persons, appear to receive meanings different from the meanings assigned them by psychologists (e. g., "deficit"). This list contains 140 words and its importance does not justify its inclusion in the present paper.

Frequency of technical words is defined as the total number of words from such a list divided by the total number of words in the report. Such a frequency count was made in each of the nine psychological reports; the distribution of values is summarized in Table 4.

"Personalizing terms" include all personal pronouns (excluding "it"), include all nouns that have a masculine or feminine natural gender (e.g., "nurse") but not common gender nouns (e.g., not "authority"), and include plural nouns with a group referent (e.g., "folks"). This definition of personalizing terms parallels the definition offered by Flesch [4] in his readability studies.

Frequency of personalizing terms is therefore defined as the total number of personalizing terms divided by the total number of words in the report. Distribution of these values for the nine reports studied is presented in Table 4.

Each of these four variables: length of report, length of sentence, frequency of technical terms, and frequency of personalizing terms, was correlated by rank-order method with the variables pertaining to utility and clarity of the report. Results of these correlations are presented in Table 5. This table also presents reliability figures for the utility and clarity variables.

[3]The author is indebted to M. Buttiglieri and R. Phillips for their analysis of these linguistic variables.

Table 4

Linguistic Properties of Nine Psychological Reports

Patient	Report length (words)	Sentence length (words)	Technical terms (per word)	Personal terms (per word)
A	730	19.7	.20	.13
B	359	15.6	.22	.08
C	467	20.3	.22	.05
D	595	20.0	.22	.07
E	254	16.9	.17	.04
F	362	17.2	.18	.05
G	430	17.9	.17	.08
H	604	18.3	.22	.06
I	382	22.5	.23	.05
Median	430	18.3	.22	.06

Table 5

Rank-Order Correlations between Clarity, Utility, and Linguistic Variables

($N = 9$ reports, each judged by two clinicians)

Linguistic variable	Number of clear, useful decisions	Number of new, clear, useful decisions
Report length	.66	.27
Sentence length	−.10	−.15
Technical terms	.64	.43
Personal terms	.36	.32
Number of clear, useful decisions	.59*	
Number of new, clear, useful decisions		.58*

* These correlations are reliability coefficients calculated by the split-half method.

A Priori Likelihood of a Decision

Because of the uniformly low rank-order correlations involving the variable, "number of new, clear, useful decisions" (Table 5), it is appropriate to question the assumption that the weighting system used in measuring this variable has any actual relationship to the process of clinical decision. Are the recommendations for therapy made by clinicians when they are given *no* information about a patient (Table 2) similar to the recommendations which they

make after obtaining information about a specific patient?

The author demonstrated in previous experiments [2] that the stating of an opinion about a person led a judge to "fixate" his opinion. As a further question about the nature of clinical "rigidity," the similarity between clinical judgments before and after reading a report should be measured.

Both these questions (the reliability of the weighting system and the rigidity of clinical recommendations) converge in the data presented in Table 6. This table compares the amount of agreement, among ten clinicians, in recommending each of the 32 treatment procedures for the "hypothetical average N-P patient," with the amount of agreement among those same clinicians in recommending the 32 treatment decisions for the specific patients whose reports they read. Thus, each treatment decision receives two agreement scores. Table 6 compares the treatment recommendations "before" and "after" reading reports.

Table 6

Correlation between Decisions Made about the "Average" Patient with Decisions Made about 9 Specific Patients

Number of clinicians making a given decision for the hypothetical average N-P patient *	Number of times same decision made regarding any specific patient after his psychological report read by a clinician †		
	0–3	4–9	10–18
0	7	2	0
1–6	3	8	3
7–10	1	2	6

$N = 32$ decisions
Chi square $= 17.4$‡
$df = 4$
Significance less than .01

* Ten clinicians made these decisions for the "hypothetical, average N-P patient" (Table 2), making a total of ten sets of decisions.

† Nine reports were each read by two clinicians; each clinician made a set of decisions for each report that he read, making a total of eighteen sets of decisions.

‡ The Pearsonian r for these data, when they are ungrouped, is .76. This is a highly significant value.

Table 6 demonstrates that the weighting system used in measuring the variable, *number of new, clear, useful decisions*, does seem to have a relationship to the process of clinical decision about what treatment should be given to *specific* patients. Whether this also demonstrates the strong influence of clinical bias upon treatment planning remains to be determined by further experimentation. The finding could also be interpreted as indicating that the information on which reports are based (i.e., psychological testing) is poorly developed in the direction of treatment planning.

It seems safe to conclude (*a*) that the weighting system has some reliability and (*b*) that the factor of clinical bias (among psychologists, at least), or stereotypy of opinion, helps determine the role of the report in the treatment-planning process. Such conclusions have interest in relation to the method of measurement described in this paper.

Conclusions Regarding the Measurement of Utility and Clarity

1. The number of clear, useful decisions implied by a psychological report can be measured by counting the number of decisions agreed upon by two clinicians who have read the report independently. This measure has a split-half reliability of .59 in the small sample of nine reports studied.

2. The number of *new*, clear, useful decisions implied by a psychological report can be measured by weighting each decision in (1) above, according to the inverse of the likelihood that such a decision could be made with reference to the average patient. This measure has a split-half reliability of .58 in the small sample studied.

3. The weighting procedure employed in (2) appears to be quite reliable in that clinicians tended to make the same therapeutic decisions after reading psychological reports that they had previously made for the "average N-P patient." This similarity between a priori and a posteriori recommendations may explain why greater relationships between the "readability" and other linguistic variables on the one hand, and, on the other hand, the variable *new, clear, useful information,* were not obtained. The operation of clinical bias or rigidity may be so great as to mask many of the effects of the psychological report.

4. Suggestive but not convincing relationships between the number of clear, useful decisions and the length and technical content of the report were found. The longer a report

(between the values of 254 and 730 words), the greater the clarity in terms of therapeutic implications; and, the more technical the terminology (independent of length), the greater the clarity.

Implications for Further Research

The above data and findings are presented only for illustrative purposes. The intent of this study is primarily methodological.

The reliabilities of the measures of utility and clarity are high enough to be promising but low enough to require further refinement of the measures before large-scale inquiry is justified.

Limitations on the general use of such measures are imposed by the fact that, logically, such measures have to be devised with reference to the institutions within which they are to be used. Variation in institutional lists of decisions and particularly in a priori weights (i.e., clinical premises) would reduce the general applicability of any one scale.

It should also be emphasized that the present method defines only one dimension among the many along which the value of a psychological report should be calculated. A report has relevance not only to gross decisions such as those listed here, but also to subtler problems which arise during protracted psychotherapy. Other purposes of the report—such as establishing in the therapist or decision maker a clear and memorable image of the patient—are not met by the present method.

Nor does the method help solve the problem of whether *valid* decisions have been suggested by the report. The broader problem of whether utility and clarity have any relation to validity remains for future research.

The most important suggestion of the present discussion is that the "action-potential" of a report should be a criterion of its usefulness; the most important implication of our data is that it is possible to measure this property. It is possible that we may uncover empirical criteria which will ultimately define the language which should be spoken by the clinician.

Received November 6, 1952.

References

1. Cronbach, L. J. A generalized psychometric theory based on information measure. (Mimeographed paper.) Urbana, Ill.: Bureau of Research and Service, Univer. of Illinois, 1952.
2. Dailey, C. A. The effects of premature conclusion upon the acquisition of understanding of a person. *J. Psychol.*, 1952, **33**, 133–152.
3. Davenport, Beverly F. The semantic validity of TAT interpretations. *J. consult. Psychol.*, 1952, **16**, 171–175.
4. Flesch, R. *The art of readable writing.* New York: Harper, 1949.
5. Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor, Mich.: Univer. of Michigan Press, 1951.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Clinical Judgment and the Cattell Infant Intelligence Scale

James J. Gallagher
Michigan State College

It is a maxim in psychology that the efficiency of clinical practice is determined by the validity of the clinical judgment of the individual psychologist. Although called upon to use this judgment every day in practice, the average psychologist has had little opportunity to check systematically the validity of his own judgment because of pressures of business and some difficult methodological problems. One of the many situations in which the clinician must use his clinical judgment is when he is asked to test infants and give an opinion as to their intelligence, oftentimes at the request of an adoption agency. Good judgment is especially needed in this situation since, for obvious reasons, test conditions are often far from ideal when infants are being evaluated. Every clinician has in his or her reports written some such comment as:

"This child did not appear to be doing his best."
"The child was so irritable that the examiner feels her performance was lower than it otherwise would have been."
"The infant appeared sluggish and sleepy and did not respond to many of the items."

Yet the opportunity of checking these statements comes only too seldom to the busy clinician. The purpose of the present study was to determine if there were any significant changes upon retest in a group of children who were judged as not doing their best upon the original infant test.

Previous Studies

The entire area of infant testing has been a neglected one in psychology. Although both the Gesell Developmental Scales [3] and the Cattell Infant Intelligence Scale [1] have been in common use for some time, little has been reported concerning their validity and reliability beyond the original presentation of the authors. A new scale designed for infants from 4-12 weeks has recently been developed at Northwestern University [4] but is not yet in general clinical use. Watson [6] has recently stated that no published study, other than that of Cattell herself, concerning the scale was known, although it was then ten years old. Since then Escalona [2] has done a study on the use of infant tests for predictive purposes. Using both the Cattell and Gesell scales she found that a group of children tested under nonoptimal conditions tended to vary more on retests than did a group of infants who were tested under what was judged to be optimal testing conditions. The measure of change was, in this case, movement in broad intellectual classifications, i.e., such as from Dull Normal to Average. Even the group tested under optimal conditions showed a marked tendency to vary one intellectual degree away from the original test in 31% of the cases. Her conclusions were that for the best predictive results the gestalt of the testing situation had to be taken into account rather than the mere test finding itself.

Method

The subjects in the present study were forty-three infants, ranging in age from 4.1 months to 24.1 months, who were seen at the Dayton Guidance Center over a period of three years. All these infants had been referred by community agencies supervising adoption cases, and all had taken the Cattell Scale at least twice. Since the general policy of the psychologists was that retests of all infants were desirable, it was possible for the writer to divide the children into two groups: those having routine retest and those who were suspected

of not doing their best on the original tests. The two psychologists who did the tests had M.A.'s in psychology plus several years of experience.

The writer, who did none of the actual testing, chose the children for the Mandatory Retest group upon the following criteria:

a. Evidence in the case history of extremely deprived environmental background up until the present examination (such as malnutrition, desertion by parents, etc.).

b. Definite statements in the psychological report that owing to poor rapport conditions (child crying — restless — not responding) the psychologist believed retesting necessary.

The children placed in the Routine Retest group were selected on the basis of the absence in the psychological reports of any comments indicating special problems existing in the test situation or in their previous environment. The placement of each case in its proper group was quite clear owing to the precise wording of the reports. A recheck of the placement of twenty-five of the cases, chosen at random and read two months after the original selection by the writer, found twenty-three of the twenty-five cases being placed in the same category as before. The mean age of Routine Retest group at the time of first testing was 8.29 months, while the mean age for the Mandatory Retest group was 10.53. This was not a statistically significant difference. The mean time interval between the first and second test for the Routine group was 7.90 months and for the Mandatory group was 7.18 months.

Table 1

Test-Retest IQ Scores on Mandatory and Routine Retest Groups

Measure	Mandatory Retest	Routine Retest
First test		
N	19	24
Mean	88.05	100.62
SD	17.80	9.18
Retest		
N	19	24
Mean	96.58*	101.25
SD	14.79	8.99
Test-Retest correlation	.77	.83†

* Difference between means significant at the 1% level of confidence.
† Corrected for restriction of range [5].

Table 2

Changes in Terman-Merrill Classifications As Result of Retest

Test	Two below	One below	Same category	One above	Two above	Three above
Routine Retest		4	16	4		
Mandatory Retest			11	7		1

The standard *t* test was used to determine the significance of difference between the pre- and posttest mean IQ's of each group. The changes in the standard Terman-Merrill [5] classifications as a result of the retest were also calculated.

Results

Table 1 shows that the Mandatory Retest group made a mean gain of 8.53 IQ points at the time of retest, and the difference between first and second mean IQ scores was significant at the 1% level of confidence. The Routine Retest group showed a gain of .53 IQ points on the retest, a difference that was not statistically significant. The test-retest correlation for the Mandatory group shows a respectable but not outstanding *r* of .77. The Routine Retest group showed a test-retest *r* of .52 but when this is corrected for restriction of range, this *r* becomes .83.

Table 2 reveals the change in Terman-Merrill classification as a result of the retest of both Routine and Mandatory Retest groups. In the Routine group the differences were evenly split, four infants dropping one classification while four moved up one classification. These results closely agreed with Escalona's findings of her optimal test group, with 31% of her cases falling in an adjacent range, as compared with 33% of the present study. In the Mandatory group it may be seen that the changes are all in the direction of a higher retest classification and one infant even changed to three classifications higher.

Conclusions

1. In this particular situation it appears that the clinical judgment of the psychologists was sound in stating as necessary further testing for those infants who did not appear to be doing their best. While these results are limited to the situation involved in the study, it is

encouraging to note the similarity of the results in this study and in Escalona's study.

2. It can be noted that although the results of the Routine Retest group show them to have a much more stable retest pattern than the Mandatory Retest group, their own reliability cannot be considered too good for practical circumstances. All doubtful cases should be seen again rather than risk damage to a prospective adoption case by too much reliance on a useful but not overly reliable instrument.

3. Evidence seems to support the conclusion that all possible information concerning not only test results but the conditions surrounding the test and full case histories are necessary for adequate competence in the prediction of infant intelligence.

Received October 30, 1952.

References

1. Cattell, Psyche. *The measurement of intelligence of infants and young children.* New York: Psychological Corporation, 1940.
2. Escalona, Sibylle. The use of infant tests for predictive purposes. *Bull. Menninger Clin.,* 1950, **14**, 117–128.
3. Gesell, A., & Amatruda, Catherine. *Developmental diagnosis.* New York: Hoeber, 1941.
4. Gilliland, A. R. *Northwestern intelligence tests. Test A. Test for infants 4–12 weeks old.* Boston: Houghton Mifflin, 1949.
5. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1950.
6. Terman, L. M., & Merrill, Maude A. *The measurement of intelligence.* Boston: Houghton Mifflin, 1937.
7. Watson, R. I. *The clinical method in psychology.* New York: Harper, 1951.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Three Factors in Parent Behavior

Maurice Lorr and Richard L. Jenkins
Veterans Administration[1]

The importance of the family and the role of the parents in the personality development of the child have been widely recognized. Much clinical evidence has accumulated with regard to the influence of parent attitudes, and to the effect of various methods of handling, on the child's emotional development. However, until recently, comparatively little systematic effort has been directed toward the identification of basic variables in parent behavior.

The Fels Parent Behavior scales developed by Champney [2] represent an effort to categorize and measure a wide range of family variables. The variables covered in the 30 rating scales were intended to measure eight basic "armchair" factors of parent behavior. The eight factors or areas describe the following: home atmosphere, contact of child and mother, babying and protectiveness, criticism and evaluation of the child, readiness of explanation, emotional relationships between parent and child, and a miscellaneous group.

Baldwin, *et al.* [1], using the rating schedule developed by Champney, secured ratings by home visitors on factors in each home in a population whose children were also being studied in other ways. Subsequently, in an effort to secure simplification of the 30 variables, Roff [3] isolated seven common factors by means of a multiple factor analysis of the intercorrelations between the Fels Parent Behavior scales obtained by Baldwin.

[1] From Veterans Administration Central Office, Washington, D.C.

Table 1

Matrix of Correlations of Primary Factors

		I	II	III	IV	V	VI	VII
Concern for child	I	1.00						
Nonreadiness of suggestion	II	−.36	1.00					
Democratic guidance	III	.37	.06	1.00				
Parent-Child harmony	IV	.38	.11	.63	1.00			
Nonpermissiveness	V	−.36	.14	−.34	−.36	1.00		
Activeness of home	VI	.03	−.11	.27	.04	.36	1.00	
Sociability of parents	VII	.23	.26	.32	.06	.22	.26	1.00

Table 2

Rotated Oblique Second-Order Factor Matrix r_{ig} and Transformation Matrix ψ_{jg}

Rating Scales		r_{ig} X	r_{ig} Y	r_{ig} Z	ψ_{jg} X	ψ_{jg} Y	ψ_{jg} Z
Concern for child	I	.781	.012	.092	1.071	.016	.126
Nonreadiness of suggestion	II	.578	−.338	−.012	−.732	.428	.015
Democratic guidance	III	.022	.770	−.001	.034	1.197	−.002
Parent-Child harmony	IV	.011	.592	−.236	.014	.735	−.293
Nonpermissiveness	V	−.068	−.448	.601	.096	.633	−.849
Activeness of home	VI	.134	−.153	.569	.174	−.198	.734
Sociability of parents	VII	.001	−.005	.440	.001	−.007	.602

Table 3
Factor Loadings of Rating Scales on Second-Order Factors

Rating Scales	*X*	*Y*	*Z*
Babying: overhelps — withholds help	.651	−.006	.054
Child-centeredness of home: child-centered — child-subordinated	.640	−.025	.078
Solicitousness for child's welfare: anxious — nonchalant	.622	.133	.035
Protectiveness: sheltering — exposing	.600	.023	.245
Intensity of contact: vigorous — inert	.576	.035	.060
Duration of contact with parent: extensive contact — brief contact	.556	.054	.019
Acceptance of child: devotion — rejection	.438	.280	.015
Readiness of suggestion: suggesting — nonsuggesting	.385	−.294	.061
Democracy of policy: democratic — dictatorial	−.039	.764	−.087
Noncoerciveness of suggestion: optional — mandatory	−.029	.735	−.033
Justification of policy: rational — arbitrary	.019	.696	.030
Nonrestrictiveness of regulations: freedom — restriction	−.294	.661	−.262
Nonemotionality toward child: objective — emotional	−.204	.625	−.211
Nonreadiness of criticism: uncritical — critical	−.263	.617	−.327
Readiness of explanation: satisfies curiosity — thwarts curiosity	.060	.586	.198
Favorableness of criticism: approval — disapproval	.199	.572	.004
Understanding of child's problems: keen — obtuse	−.027	.490	.256
Clarity of policy: clear — vague	.001	.461	.262
Nondisciplinary friction: concordant — contentious	.080	.441	−.145
Affectionate toward child: affectionate — hostile	.321	.387	−.130
Rapport with child: close rapport — isolation	.312	.347	−.128
Nondiscord in home: harmony — conflict	−.019	.328	.124
Effectiveness of policy: successful — unsuccessful	.108	.299	−.063
Readiness of enforcement: vigilant — lax	.016	−.318	.592
Severity of actual penalties: severe — mild	−.079	−.425	.473
Accelerational attempt (pushing): acceleratory — retardatory	.123	.229	.427
Coordination of household: coordinated — chaotic	.079	.074	.407
Activeness of home: active — inactive	.153	−.138	.394
Adjustment of home: well-adjusted — maladjusted	.030	.098	.369
Sociability of family: expansive — reclusive	−.056	−.044	.288

The factors isolated were interpreted and described as concern for the child, democratic guidance, permissiveness, parent-child harmony, sociability adjustment of parents, activeness of home, and nonreadiness of suggestion. The correlations between these primary factors as determined by Roff [3] are shown in Table 1. Examination of the table reveals substantial correlations between these factors. It appeared that an analysis of these correlations should reveal one or more underlying second-order factors of a broad character which should be interesting in themselves and useful for the conceptualization of parent-child relations. Accordingly, the intercorrelations in Table 1 were subjected to a centroid analysis and three orthogonal factors were extracted and rotated to oblique simple structure.

Table 2 presents the loadings of the first-order factors on the second-order factors. It also gives the transformation matrix for computing the factor loadings of the original 30 rating scales on the second-order factors [4].

It will be noted that the significant loadings on Factor *X* are from concern for the child and nonreadiness of suggestion.

The variables with the largest loadings on factor *Y* are democratic guidance, parent-child harmony, and permissiveness. Nonreadiness of suggestion also has a sizable negative loading.

The heaviest loading on factor *Z* is a loading from nonpermissiveness. Activeness of the home and sociability of the home also show large loadings.

The character of these second-order factors becomes much clearer when we refer to Table 3, which presents the correlations of the original 30 family characteristics with the second-order factors. The largest correlations with factor *X* are babying, child-centeredness of the

home, solicitousness for the child's welfare, protectiveness, intensity of contact, duration of contact, and acceptance of the child; devotion rather than rejection. It would appear that this factor might be characterized as the satisfaction and encouragement of the child's emotional dependence on the parent. We will call it dependence-encouraging.

The second factor, *Y*, is best described by democracy of policy, noncoerciveness of suggestion, nonrestrictiveness, nonemotionality toward the child, nonreadiness of criticism, readiness of explanation, favorableness of criticism, understanding of the child's problems, clarity of policy, and the absence of disciplinary friction. This is clearly a factor of democracy and democratic values in child training in contrast with undemocratic methods. This factor will be called democracy of child training.

The third factor, *Z*, is characterized by vigilance of enforcement, severity of actual penalties, accelerational attempt (pushing), and coordination of household. Activeness of the home, the adjustment of the home, and the sociability of the family show smaller correlations. This is a factor of organization and effectiveness of control in the home. It ranges from a strict orderliness at one extreme to a chaotic disorder at the other.

It would appear from this analysis that, if we are to seek in clinical work to conceptualize parent-child relations in shorthand fashion, the three most economical questions we may ask ourselves are:

1. How far does this home sustain and encourage dependence and how far does it deny satisfaction to dependence?
2. How far do its methods of child training reflect democratic practices and values, or to what extent are they authoritarian and undemocratic?
3. To what extent is there a strict orderliness in the home or to what extent is the home lax and unorganized?

We must recognize, of course, that these three questions, based on the second-order factors, would not yield us all the information we would obtain from ratings on the seven first-order factors, any more than the latter would give us all the information we would obtain from ratings on the 30 schedules from which they were derived. Busy clinicians, however, can hardly afford routinely to make ratings on the 30 schedules, unless they are pursuing this as a research project. We submit that in terms of our present knowledge the three foregoing questions embody more of the significant elements of the larger study than any other equally economical condensation.

The intercorrelations between the three second-order factors were also computed. The correlation of $-.43$ between factors *X* and *Y* imply that the home with democracy of policy tends not to be dependence-encouraging. The correlation of $-.47$ between factors *Y* and *Z* suggests that a democratic policy is related to a home that is less organized and less effective in the enforcement of its policies than an authoritarian home may be. Factors *X* and *Z* are independent, indicating that there is no significant relation between the encouragement of dependence on the one hand and effectiveness of control in the home on the other.

Summary

A second-order factor analysis of the interrelations of seven first-order factors derived by Roff from 30 Fels Parent Behavior Scales indicates that the most economical and significant condensation of the significant relations tapped by these scales can be represented in the following three questions:

1. How far does this home sustain and encourage dependence and how far does it deny satisfaction to dependence?
2. How far do its methods of child training reflect democratic practices and values, or to what extent are they authoritarian and undemocratic?
3. To what extent is there a strict orderliness in the home or to what extent is the home lax and unorganized?

Received November 6, 1952.

References

1. Baldwin, A. L., Kalhorn, Joan., & Breese, Fay H. Patterns of parent behavior. *Psychol. Monogr.*, 1945, **58**, No. 3 (Whole No. 268).
2. Champney, H. The variables of parent behavior. *J. abnorm. soc. Psychol.*, 1941, **36**, 525–542.
3. Roff, M. A factorial study of the Fels parent behavior scales. *Child Develpm.*, 1949, **20**, 29–44.
4. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

Follow-up Study of Ten-Year-Old Boys with Police Records

William W. Wattenberg and Frank Quiroz
Wayne University

The present study is an attempt to check more directly some inferences drawn from the findings of an investigation reported by Wattenberg [3]. Data relating to eleven-year-old boys who had charges filed against them with the Detroit police were collected. These indicated that many of the boys were victims of neither severe personality disorganization nor bad environmental pressures. Rather, their misconduct presumably could be accounted for as resulting from poor judgment, weak internal controls, or other immaturities which they could be expected to outgrow.

There are sound theoretical grounds for believing this could be the case. Several recent writers [1, 2] on the psychology of preadolescence maintain that this period of development is accompanied in American culture by personality attributes, temporary in nature, which might lead a normally developing youngster to engage in adult-prohibited actions. These would not necessarily be symptoms of serious personality maladjustments or character defect. Although such actions might bring a youngster to the attention of the police, they would not be predictive of delinquency in subsequent years.

Two specific trends attributed to preadolescence as a growth phase are: (*a*) a disorganization of the childish personality structure and (*b*) a strong attraction to peer groups which engage in a species of guerrilla warfare against adults. Either trend, if it exists, could easily bring a youngster into collision with the forces of law and order. Both, presumably being temporary, eventually wane. Delinquency resulting from either or both would therefore be expected to vanish.

If there is any truth in these speculations, it has very practical implications. At the present time, clinics serving youth are badly overloaded in most areas. A good portion of their load consists of children referred by courts or brought by parents who have been disturbed by juvenile misconduct. Now, if in this load there is a large proportion of boys and girls who would have recovered even if nothing had been done for them, the overworked clinics are wasting precious time and even more precious skill. If such should be the case, then an obvious next step would be to search intensively for some reasonably effective means of discovering which preadolescents are genuinely in need of therapy. If this could be done, then clinics could concentrate on those youngsters who would become delinquent or mentally ill if they received no treatment.

From the viewpoint of scientific research designed to evaluate methods of psychotherapy the possibility we have been discussing has grave implications. If among the populations receiving treatment there is a large proportion who would have improved in the course of time without therapy, statistics based on percentages of "success" become meaningless. As a minimum precaution research workers would have to make allowances for the existence of this factor.

The study [3] mentioned contained indirect evidence that preadolescent boys in trouble with the police included a large proportion who were relatively normal youngsters. However, the data were neither gathered nor ordered in such a way as to give an estimate of how large the "normal" group might be nor to verify the hunch that they would recover "spontaneously" from their current misbehavior. Accordingly, it was decided to take a sample of ten-year-olds who had police records and to make a follow-up study to discover how many of them did continue to be delinquent. In addition, an effort

was made to discover what available facts gathered at the time of their first police contact distinguished between those who continued to get in trouble and those who later had no police records for a twelve-month period.

For the purposes of this study, we accepted the following hypotheses as representing the points to be tested:

Hypothesis No. 1: In a group of ten-year-olds with police records, only a small fraction will be found two years later to be continuing to act as delinquents.

Hypothesis No. 2: The group which continues to act as delinquents will be distinguished by having greater pressures in the form of bad family situations, poor school relations, and bad neighborhood influences.

Procedure

Use was made of the very comprehensive records kept by the Youth Bureau of the Detroit Police Department.[1] The records for 1948 were searched, and all those pertaining to boys ten years old at the time of their first contact were segregated. These records were then placed in order by the serial number assigned to the boys. Next, the records of all boys contacted by the police during 1950 were placed in serial number order. The two lists were checked and the original records of all boys among our original ten-year-olds who had police contacts in 1950 were set aside. We termed these boys "repeaters." The nonrepeaters, then, were boys who after a period of at least one year (but averaging eighteen months) had gone for a full twelve months without any complaint against them having been made to the police.

The year 1948 was chosen because at the time the routine handling of ten-year-olds was either to release them to their parents with a warning or to have them come in to visit police officers weekly for a series of check-up conferences. Only where the offense was very serious did the police officers file a petition with the Juvenile Court. At the time, the clinic attached to the court had a heavy backlog of cases. With rare exceptions, notably in the instance of sex offenses, the first offenders were put on probation without being seen by the clinic. Therefore, it is safe to say that only a tiny fraction of the 1948 ten-year-olds had received systematic psychotherapy.

To test the second hypothesis, the repeaters and nonrepeaters were compared on 50 items of information reported in the Youth Bureau "history sheet," filled out by selected police officers after a visit to the home, an interview with the boy, and conferences with school personnel. For each item a tabulation was prepared and tested for statistical significance by the chi-square technique. In these comparisons only the records made at the time of the first police contact were used.

Findings

In all, the 1948 files contained records of first contact with 207 ten-year-olds. In 1950, only 43 of these had any contact with the police. Thus, close to 80 per cent of the group had gone a full year without any reported relapse into actionable misconduct. These figures furnish strong support to the first hypothesis.

In the comparisons between repeaters and nonrepeaters, the null hypothesis could be rejected at the 1 per cent level of confidence for three tables; at the 2 per cent level, for an additional tabulation; and at the 5 per cent level, for four more. The chance expectation would call for 2.5 tabulations to meet this criterion. Thus, the number of statistically significant differences was 200 per cent greater than should have occurred by chance alone. However, these results are clouded by the fact that although classifications were combined, in most instances of apparent statistical significance, the chance expectation in the critical cells fell below five. For what they may be worth, however, the results will be reported.

Of the apparently significant tables, two concerned the parents. As shown in Table 1, among the few boys for whom parental alcoholism was reported, almost half were listed among the boys who were still running up police records two years later. Also, as shown in Table 2, the majority of boys whose parents were regarded as uncooperative by the police, were also repeaters. As indicated by the column, "Contribution to chi square," in both these tables, the cells yielding the highest contribution to the chi-square table had chance expectations below five. Thus, the results can

[1]The authors wish to express thanks to Senior Inspector Sanford Shoults, Inspector Ralph Baker, and Lieutenant Francis Davey for their contribution to this research. They were responsible for maintaining the records used and were helpful in making the basic tabulations possible.

be considered suggestive, but not conclusive.

Table 1
Homes in Which Excessive Drinking Led to Abuse of Boy

Excessive drinking	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
Present	7	3	4	4.67
Not present	197	40	157	0.03
Not stated	3	0	3	1.50
Total	207	43	164	6.20

$\chi^2 = 6.60$; $df = 2$; $0.05 > p > 0.02$

Table 2
Attitude of Parents toward Police Officer

Attitude	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
Cooperative	190	34	156	0.81
All others	17	9	8	8.17
Total	207	43	164	8.99

$\chi^2 = 8.99$; $df = 1$; $p < 0.01$

Two of the significant tables related to school. As shown in Table 3, almost half of the boys who, at ten, were receiving failing grades in school continued to get in trouble. Also, as shown in Table 4, a similar proportion of boys who were in special classes (in Detroit this means classes for mentally retarded youngsters) became repeaters. Here again, however, the statistical significance of the results is clouded by the small numbers involved.

Table 3
School Marks

School marks	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
Good	27	3	24	1.83
Average	119	24	95	0.05
Poor	45	9	36	0.00
Failing	16	7	9	6.56
Total	207	43	164	8.44

$\chi^2 = 8.44$; $df = 3$; $0.05 > p > 0.02$

Table 4
Grade in School

Grade	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
3rd or lower	41	9	32	0.16
4th	69	14	55	0.00
5th or higher	82	13	69	1.20
Special class	15	7	8	6.67
Total	207	43	164	8.03

$\chi^2 = 8.03$; $df = 3$; $0.05 > p > 0.02$

The statistical fault does not pertain to the results reported in Table 5. A high proportion of the repeaters had two or more brothers. The statistical significance is clear. Interestingly, there was no such relationship with number of sisters. Another index to living conditions, the type of dwelling unit occupied by the boy's family, as shown in Table 6, was also related to repeating. Those boys who lived in apartments or rooming houses were prone to become repeaters. The chance expectations in the key cells of this table were amply large to warrant a rejection of the null hypothesis without cavil. A third item related to family economic conditions was the need for a boy to employ part of his funds for clothes, maintenance, or similar uses. Among ten-year-olds this is rare, as indicated in Table 7. It was more prevalent for those boys who later became repeaters, but the numbers concerned were too small to warrant accepting the chi square at its face value.

The last item to be reported has a strange history. The present study is one in a series begun at the request of police officials. Throughout there has been much give and take on topics to be investi-

Table 5
Number of Brothers

Number of brothers	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
0	61	10	51	0.88
1	66	6	60	5.80
2	39	14	25	5.66
3 or more	41	13	28	3.89
Total	207	43	164	16.23

$\chi^2 = 16.23$; $df = 3$; $p < 0.01$

Table 6
Type of Dwelling Unit

Type of unit	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
Single-family house	93	22	71	0.59
Flat	81	10	71	3.65
Apartment or rooming house	30	11	19	5.21
Not stated	3	0	3	1.50
Total	207	43	164	10.95

$\chi^2 = 10.95$; $df = 3$; $0.02 > p > 0.01$

Table 7
Boy's Disposition of Money

Disposition	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
All spent for entertainment	193	37	156	0.31
All or part required for clothes, maintenance, etc.	14	6	8	3.82
Total	207	43	164	4.13

$\chi^2 = 4.13$; $df = 1$; $0.05 > p > 0.02$

gated. When the "history sheets" from which our data are taken were designed, the police seemed anxious to include a judgment as to the "type" of a boy. One of the "types" they listed was "Peck's Bad Boy," by which they meant a youngster who frequently did things which angered the adults of the neighborhood. Although hardly the sort of "type" that had meaning to psychologists, it was included as a token of our willingness to accept police estimates of what was important. It must be confessed that it would be more accurate to say the item was kept in to humor the police. The fact is that in all of the studies we have made, this item turns out to be highly predictive. Whatever it is in a boy that makes an experienced policeman check him as a "Peck's Bad Boy," it involves recognition of behavior trends which do indicate an important and persistent propensity toward delinquency. As shown in Table 8 this item once again proved predictive. The numbers involved were sufficiently large to warrant rejecting the null hypothesis at the 1 per cent level of confidence without reservation. Remember in this case the judgment was recorded at the time of first contact and proved predictive of conduct two years later.

Table 8
"Type" of Boy

"Type"	Total	Number Repeaters	Number Non-repeaters	Contribution to chi square
Normal	143	21	122	3.42
"Peck's Bad Boy"	52	20	32	9.34
All others	12	2	10	0.44
Total	207	43	164	13.20

$\chi^2 = 13.20$; $df = 2$; $p < 0.01$

Although detailed examination of the tables did bring out some predictive items, these did not fall into the pattern required to support our second hypothesis. The findings in relation to home and school conditions were inconclusive. Clear-cut results pointed to frequent migration (as implied by living in rooming houses), the presence of many brothers, and the existence of strong adult-annoying characteristics as differentiating the repeaters from the nonrepeaters.

Discussion

The major finding of this study is that only a small proportion of ten-year-olds against whom complaints are registered with the police continue to get in trouble. Before accepting this as bearing on the point that preadolescent misconduct may be symptomatic of conditions which are self-correcting or which disappear as a result of maturing, we must weigh the possibility that many of these boys lacked police contacts because they had moved from Detroit or that they may have received therapy.

Three facts would seem to indicate that migration from Detroit was not a significant factor. First, the rooming house occupants, the group most likely to be on the move, were more heavily represented among the repeaters than the nonrepeaters. Second, the period, 1948 to 1950, was a relatively stable one during which there was little emigration from the city. Third, the fact that there were statistically significant differences not related to economic circumstance differentiating between repeaters and nonrepeaters would nullify any contention that the two groups were alike except that the nonrepeaters happened to leave the police jurisdiction. We can safely discount the argument that the ten-year-olds had left town in any major proportion.

As to the possibility that they had received therapy, here we cannot be as confident. The indications that this was not the case are indirect. As previously mentioned, referral to the court clinic in 1948 tended to be reserved for unusual cases, especially those having sex offenses. Our data did include the offense with which the boys were charged. On this count there was no significant difference beween the repeaters and the nonrepeaters. However, there is no way of knowing exactly what each family did as a consequence of a boy getting in trouble. We do know that during this period a special arrangement existed whereby the police could get speedy action on any referrals to social agencies and that this machinery was used for only four cases. As far as the police records show, there was almost no use made of the social and clinical resources of Detroit for the boys with whom we are concerned. In general,

these services have been brought into the picture for the older youngsters who have longer records and whose families have given up hope of doing anything on their own. For the 20 per cent of the ten-year-olds who have become repeaters, this is most unfortunate.

Summary

A follow-up study was made of 207 boys who as ten-year-olds in 1948 had been the subjects of complaints to the Detroit police. Only 43 of this group had been in trouble with the police during 1950. The items recorded at the time of the first contact which proved predictive of repeating were (*a*) having two or more brothers, (*b*) living in an apartment or rooming house, and (*c*) having a reputation as a "Peck's Bad Boy."

Received December 30, 1952.

References

1. Blair, A. W., & Burton, W. H. *Growth and development of the preadolescent.* New York: Appleton-Century-Crofts, 1951.
2. Redl, F. Preadolescents—What makes them tick? *Child Study,* 1944, **21**, 44-48.
3. Wattenberg, W. W. Factors associated with repeating among preadolescent "delinquents." *J. genet. Psychol.,* in press.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

A Note on the K Scale of the Minnesota Multiphasic Personality Inventory

Anders Sweetland and Herbert Quay
Florida State University

The Minnesota Multiphasic Personality Inventory (MMPI), designed by Hathaway and McKinley [3], is a 560-item inventory comprised of several scales designed to measure in the areas of: hypochondriasis (*Hs*), depression (*D*), hysteria (*Hy*), psychopathic deviate (*Pd*), masculinity-femininity (*Mf*), psychasthenia (*Pt*), paranoia (*Pa*), schizophrenia (*Sc*), and hypomania (*Ma*). It also has four "validity" scales, designed to measure test-taking attitudes. Among these is the *K* scale [5]: a measure of "plus-getting" (a tendency to get poor MMPI scores) and "defensiveness" (a tendency to get good scores).

Despite the recommendations of several investigators [1, 7, 9], most clinicians will still venture profile diagnosis of the MMPI. Although the evidence indicates that the profile should be used mainly to differentiate between the neurotic, the psychotic, and the normal, we hazard that few psychologists will be able to avoid searching the profile for clues about a client. Since profile analysis seems to be inevitable, it seems axiomatic that it should be as adequate as possible.

Clinicians tend to approach the MMPI profile by one of two avenues: (*a*) as a nosological aid, or (*b*) as a projective technique. The latter generally begins with a careful consideration of the test items. With these well in mind the various scores become a frame of reference rather than a measure of hypochondriasis or hysteria. It is presumed that the empathetic counselor, knowing the items well, can "slip on the garb of the profile" and view the world through the eyes of the client. This, patently, is not rigorous psychology, and few pretend that it is. On the other hand, profile analysis of this kind can be uncannily accurate.

Profile analysis of this second type is not as esoteric as it sounds. The clinician thoroughly familiarizes himself with items in the various subscales. A high score tells him that a particular set of items was answered in a given manner. A low score indicates the opposite. The job of learning the items is facilitated by the fact that most of them are about what is expected. There are exceptions, but in doubtful cases one can refer to the aberrantly scored items.

The *K* scale, however, does not lend itself readily to this practice. The items fall into no apparent pattern. If they suggest anything, they seem to deny having pathological symptoms. As a consequence, the clinician uses the scale for what it was designed: as a control of test-taking attitude.

The possibility that the *K* scale might measure something other than test-taking attitude was suggested by its fluctuation during an experiment with hypnotic dreams [8]. The results indicated that symbolized dreams were produced by well-adjusted people and not by maladjusted people. The measures of adjustment were the MMPI and the Social Security-Insecurity Test (SI) of Maslow, Hirsh, Stein, and Honigman [4]. The SI is a 75-item inventory designed to measure in the area of social security.

The correlations between the adjustment scales of the MMPI and the number of symbolized dreams were low (Circa —.30) but all were negative. In contrast the correlation between *K* scores and symbolization was high and positive (.72) as was the correlation between the SI and symbolization (.57). The situation was that *K* correlated high with dream symbolization and the latter was produced only by well-adjusted people.

The dream experiment was repeated under slightly different conditions. The statistical re-

sults were essentially the same: high correlation between *K* and SI scores when paired with symbolization and negative correlations between symbolization and the adjustment scales of the MMPI. The relationships were evidently not due to chance behavior of the data. Apparently *K* was a measure of adjustment.

As a check, the *K* scale and the SI were given to seventy-five college students. The resulting correlation was 0.68. A second sample of ninety-one college students yielded a correlation of 0.74.[1] It would seem from this that the *K* scale measures in much the same area as the SI.

Two places in the literature suggest similar "misbehavior" of the *K* scores. The first was in the work of Gough [2], whose data suggest that when faking is attempted, *K* scores are depressed in a marked fashion although *K* scores of psychiatric groups ordinarily are slightly below average. (This was inferred from his data. Faked profiles show *F*—*K* scores of 14 to 41 while bona fide scores from normal and psychiatric groups range from—3 to —9.)

Gough's data, conforming with that offered originally by Meehl and Hathaway [5], also indicate that college students are characterized by elevated *K* scores. His average *K* is approximately 18 for this group (again inferred from his data), while *K* for average adults is approximately 13.

A factor analysis of the MMPI by Wheeler, Little, and Lehner [9] provides another opportunity to consider the *K* factor. Like Meehl and Hathaway they find that *K* before factoring correlates negatively with all the subscales except *Hy*. Similarly, they find a marked negative relationship when correlating *K* with *Pt* and *Sc*. (The range is from −.55 to −.67). The inference seems to be that the *K* scale measures something in common with the *Hy* scales, but something opposite the *Pt* and *Sc* scales.

In the factor analysis, *K* again appears to behave unusually. With the exception of the second factor extracted from the normal group, *K* never appears in the same factor as *Hy*, and it is the only scale that yields high negative loadings. Further, it tends to be significant only in the first factor extracted, a factor which has high positive loading of *Pt* and *Sc*.

Our interpretation differs somewhat from that of Wheeler *et al*. They interpret *K* as representing ego "defense." We believe that a more logical interpretation might be made by considering *K* as being one end of a continuum with *Pt* and *Sc* at the opposite end. This is based on the fact that when *K* and the other two scales appear in the same factor, the *K* loadings range from −.60 to −.70 while the *Pt* and *Sc* loadings are .90 and higher. From this we venture that what is measured by the *K* scale is the opposite of that measured by the *Pt* and *Sc* scales. We owe Wheeler, Little, and Lehner an apology for selecting this limited aspect of their excellent study. The interested reader is urged to consult the original article.

One further clue is provided. In a follow-up study [6], the hypnotic dreams were rated as being "extrapunitive," "intrapunitive," "impunitive," and "no feeling." The correlations between these and the *K* scores were: extrapunitive, −.81; intrapunitive, −.33; impunitive, —.20; and no-feeling, +.61. The no-feeling dreams were produced chiefly by the well-adjusted subjects. The results suggest that *K* tends to be opposite from extrapunitiveness and closely associated with good adjustment. Although the two highest correlations are reliable at the 1 per cent level, the interpretation is made with caution because of the small sample: dreams = 405, subjects = 16.

Summarizing these bits of information we might say that normal *K* scores seem to be associated with: (*a*) social security; (*b*) emotional adjustment; (*c*) a tendency to symbolize hypnotic dreams; (*d*) being a college student; (*e*) nonfaked profiles; (*f*) the opposite of what is measured by the *Pt* and *Sc* scales; (*g*) nonextrapunitiveness.

If we were to characterize the person described by this summary, he would appear to be a therapeutic ideal: a person who is comfortable in his social contacts, above average in intelligence, and little bothered with doubts and indecisions. He seems to have an ability to live each day as it comes and to "sleep well o' nights." The indication is that this is a core part of his personality and not a façade that he erects to enable him to meet the day.

[1] The correlation between the SI and the social IF scale of the MMPI was .77, the correlation between *K* and the MMPI SIF scale .48.

In view of the above, again emphasizing that the data are tenuous, it is our feeling that *K* measures not only the degree of "defensiveness" or "plus-getting" with which a person takes the MMPI, but that it may also be a measure of healthy emotional adjustment.

Summary

The possibility that the *K* scale of the MMPI might measure something in addition to test-taking attitude was discussed. The suggestion was made that K scores, when within normal ranges, may possibly be a measure of personality integration.

Received December 8, 1952.

References

1. Benton, A. L., & Probst, Kathryn A. A comparison of psychiatric ratings with Minnesota Multiphasic Personality Inventory scores. *J. abnorm. soc. Psychol.*, 1946, **41**, 75–78.
2. Gough, H. G. The *F* minus *K* dissimulation index for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1950, **14**, 408–413.
3. Hathaway, S. R., & McKinley, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. New York: Psychological Corp., 1943.
4. Maslow, A. H., Hirsh, E., Stein, Marcella, & Honigman, Irma. A clinically derived test for measuring psychological security-insecurity. *J. genet. Psychol.*, 1945, **33**, 21-41.
5. Meehl, P. E., & Hathaway, S. R. The *K* factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525-564.
6. Quay, H. C. Emotions in hypnotic dreams. Unpublished master's thesis, Florida State Univer., 1952.
7. Schmidt, H. O. Test profiles as a diagnostic aid: The Minnesota Multiphasic Inventory. *J. appl. Psychol.*, 1945, **29**, 115-131.
8. Sweetland, A., & Quay, H. An experimental investigation of the hypnotic dream. *J. abnorm. soc. Psychol.*, 1952, **47**, 678–682.
9. Wheeler, W. M., Little K. B., & Lehner, G. F. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134–141.

Journal of Consulting Psychology
Vol. 17, No. 4, 1953

The Archimedes Negative Aftereffect as an Indication of Memory Impairment

Lloyd S. Standlee

Institute of Educational Research, Indiana University

Visual fixation upon a slowly rotating Archimedes spiral produces the illusion of expansion or contraction. Which illusion is experienced depends upon the direction of rotation. The opposite illusion (negative aftereffect) is experienced when the gaze is transferred to any other fixed object. With clinical judgment as the criterion of memory impairment, Freeman and Josey report " . . . a strong probability of correlation between the negative aftereffect of the Plateau [Archimedes] spiral and the impairment of memory" [1, p. 797].

The present experiment utilized a more objective criterion to further test the possibility of this relationship between a visual phenomenon and memory ability. Twenty-five "psychotics" and 16 "normals" were tested on ability to experience the negative aftereffect and on the Wechsler Memory Scale [2]. The psychotics were tested again approximately eight hours after the administration of electric shock.

All of the 25 psychotics experienced the negative aftereffect in the first testing session. Twenty-three experienced it for both the expanding and the contracting illusion in the second testing session. One experienced it for the expanding illusion only; one for the contracting illusion only. All of the 16 normals experienced the negative aftereffect for both the expanding and the contracting illusion.

The psychotics averaged 71 MQ before shock and 75 MQ after shock. The gain of 4 MQ points is significant at the 1% level of confidence (t test on individual MQ gains). The normals averaged 120 MQ points. Combining the two groups, the MQ's ranged from below 48 to above 143—the upper and lower limits of the Wechsler Memory Scale.

Thus, Freeman and Josey's "strong probability of correlation" between the visual phenomenon and memory ability is not confirmed. All but two subjects were perfect in their ability to experience the negative aftereffect. In memory ability, however, the subjects were distributed fairly evenly through the entire range of Memory Quotients on the Wechsler Memory Scale. In other words, memory impairment to the extent of below 48 MQ does not prevent a person from experiencing the negative aftereffect.

The second testing session was included to determine whether memory ability would be impaired following electric shock. If the Wechsler performance and ability to experience the negative aftereffect were to drop, an additional measure of memory impairment would be available. Such was not the case, however. The MQ's improved significantly (in the statistical sense), and the visual phenomenon was not affected. Though the improvement in MQ's was statistically significant, it was confounded with practice effect. Therefore, it cannot be concluded that memory ability was improved by electric shock. It can be said, however, that memory ability was not impaired.

It is concluded (a) that electric shock does not impair memory ability; (b) that electric shock does not impair ability to experience the Archimedes negative aftereffect; and (c) that there is no relationship between memory ability and ability to experience the Archimedes negative aftereffect.

Received April 6, 1953.
Early publication.

References

1. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, **62**, 794–797.
2. Wechsler, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, **19**, 87–95.

NEW BOOKS AND TESTS

Books

Baker, Harry J. *Introduction to exceptional children.* (Rev. Ed.) New York: Macmillan, 1953. Pp. xvi + 500. $5.00.

A college textbook planned for the training of school personnel. Exceptional children are often a complicating problem to the smooth running of the classroom. Baker deals effectively with the problems of unusual or handicapped children and seeks to give a teacher an awareness of their educational needs. The present edition is a revision of the original publication of 1944. There have been only a few minor changes in content, but the bibliography and index have been revised and brought up to date. — *B. M. L.*

Blum, Gerald S. *Psychoanalytic theories of personality.* New York: McGraw-Hill, 1953. Pp. xviii + 219. $3.75.

In so modestly brief a volume, it is something of a surprise to find such a comprehensive survey of psychoanalysis. All shades of analytically-oriented theory are represented—Freud and the orthodox—the early deviants, Adler, Jung, and Rank—the neo-analysts, including Fromm, Horney, Sullivan, and others. The text proper, organized in terms of developmental stages, is a clear and unbiased statement of the analysts' theories. All are represented fairly, and the author does not impose his preferences. In his "notes," which occupy a third of the volume, Blum permits himself the luxury of criticism and opinion, and brings to bear much of the relevant evidence from research. He ably supports his thesis that psychoanalytic views provide the most promising present basis for a theory of personality, but that these views need to be sorted out, evaluated by research, and progressively refined. The book is a valuable reference for all interested in psychoanalysis—as who is not, these days—but especially for those disposed to approach it with a research orientation. —*L. F. S.*

Note: The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

Brower, Daniel, & Abt, Lawrence E. (Eds.) *Progress in clinical psychology.* Vol. 1, Secs. 1 & 2. New York: Grune & Stratton, 1952. Pp. xi, xxiii + 564. 2 vols. $5.75, $5.00.

After a cogent introductory chapter by Abt, the rest of this forty-two chapter, two-volume work is contributed by 38 different authors, most of them clinical psychologists. In the first volume the aim has been "to provide as complete a coverage as possible for the past six years in clinical psychology and to point up, in the process, as many stimuli as possible to further thinking and research." The writing is somewhat uneven, but the literature on the well-known intelligence and projective tests is well, if not exhaustively, covered. The following section on psychotherapy includes separate chapters on client-centered counseling, psychoanalytic theory and practice, play techniques, spontaneous art in therapy, and neurosis as a learning phenomenon. Volume 2 includes a section on developmental processes from infancy to old age, one on applications of clinical psychology to special areas (industry, rehabilitation, crime, etc.), and a section on approaches to clinical psychology which includes chapters on psychosurgery, cultural anthropology, social psychology, statistical methods, and P-technique. The volumes conclude with a thoughtful discussion of clinical psychology as a profession by Victor Raimy. The authors state their intention to issue follow-up volumes, similarly designed, every two or three years. This will necessarily cover much of the material surveyed yearly, but inevitably much more briefly, by the *Annual Reviews.*—*A. R.*

Cabot, Hugh, & Kahl, Joseph A. *Human relations: concepts and cases in concrete social science.* Cambridge: Harvard Univer. Press, 1953. Pp. xxi + 333, viii + 273. 2 vols. $4.75, $4.25.

Human Relations is the outgrowth of an undergraduate course in social science taught by the authors at Harvard. Volume I supposedly presents the concepts to enrich the cases presented in Volume II and to enable the student or reader to deal with the cases, preferably, one gathers, with the aid of an instructor. If Volume I were called "Readings in the Social Sciences," rather than "Human Relations," it would be more appropriately labeled. Per-

haps 90 per cent of the first volume consists of extensive excerpts, frequently complete chapters, from the works of well-known sociologists, psychologists, psychiatrists, philosophers, logicians, semanticists, and anthropologists. The authors usually supply brief introductions to these extensive quotations. The approach purports to be clinical, but it is anything but that. This volume is an excellent collection of first-rate writing in a variety of social science fields, but does not add up to a satisfactory treatment of the subject of human relations. Volume II consists of 33 cases portraying a great variety of situations in which "solutions" or "treatment programs" are not offered. This makes them useful for class discussion. The treatment of the cases is "common sense" rather than clinical. This is not said disparagingly since, to many, such an approach to human relations is preferred. It is questioned in this publication only because the source material presented in Volume I is largely from profound scholars and social scientists. The incongruity is striking, since the two volumes, which are intended to be used together, are practically unrelated.—*M. K.*

Hamrin, S. A. *Initiating and administering guidance services.* Bloomington, Ill.: McKnight & McKnight, 1953. Pp. xi + 220. $3.00.

The author states that his object is to aid and promote understanding and encouragement of guidance activities by school administrators. The topics discussed include the initiation of a program, in-service education, and public relations. In addition, chapters are devoted to descriptions of the traditional services such as orientation, individual study, counseling, occupational information, group activity, placement, and follow-up. There are many excerpts from unprinted reports on existing programs by school practitioners. It contains little content of a psychological nature. Although the author points out how few guidance workers have formal training, there is not enough in the volume to indicate that the best guidance means more understanding of human motivation, frustration, and conflict resolution both in client and counselor. Such an emphasis seems, to the reviewer, appropriate even in a book emphasizing administration.—*F. McK.*

Hoyles, J. Arthur. *The treatment of the young delinquent.* New York: Philosophical Library, 1952. Pp. xi + 274. $4.75.

A British clergyman undertakes "a careful examination of the lessons of history and the findings of modern science in the light of Christian principles." It leads him to the conclusion that "there is an alternative to punishment."—*E. L. K.*

Patterson, C. H. *The Wechsler-Bellevue Scales: A guide for counselors.* Springfield, Ill.: Charles C. Thomas, 1953. Pp. xiii + 146. $3.75.

This monograph is a thorough and competent review of the research literature about the Wechsler-Bellevue, with a bibliography of 250 entries. The presentation is well organized, and each major section ends with an integrative summary of the conclusions justified by the evidence. It is indeed a "guide for counselors" in the sense of displaying the values of the test and also its limitations.—*L. F. S.*

Trenaman, Joseph. *Out of step.* New York: Philosophical Library, 1952. Pp. xx + 223. $4.75.

A fascinating account of an attempt to rehabilitate young but seriously delinquent soldiers, carried out by the British army during the war period. A sample of 200 cases was selected for intensive study and compared with nondelinquent soldiers of the same age with respect to a large number of physical, psychological, and sociological variables. Although only one-third of the cases had a history of civilian offenses, the great majority appeared to have deep-seated problems likely to lead to conflict with either civilian or service authority. Most of the book is concerned with an analysis of the differences between delinquent and nondelinquent soldiers, but it includes many details of the remedial-therapeutic program, of four- to six-months duration, after which slightly over half of the cases made a satisfactory adjustment to army life.—*E. L. K.*

Weinland, James D., & Gross, Margaret V. *Personnel interviewing.* New York: Ronald Press, 1952. Pp. vii + 416. $6.00.

A general text dealing with interviewing as practiced in business and industry. It is apparently intended for students with little or no previous training in psychology, since many of the chapters read like an elementary text in psychology. For example, one chapter is devoted to individual differences and another to a discussion of psychological tests. Other chapters, however, discuss almost every technique of interviewing, and specific types of interviews used in the industrial setting. The only bibliographical references are those included in an occasional footnote.—*E. L. K.*

Tests

Grassi, Joseph R. *The Grassi Block Substitution Test for Measuring Organic Brain Pathology.* Manual, pp. 75 ($3.00). Springfield, Ill.: Charles C Thomas, 1953.

The *Grassi Block Test* is an ingenious modifi-

cation of the Kohs technique which involves two novelties, the use of three-dimensional reproduction, and the observation of ability to "shift" to a new color pattern. Each of five models, constructed from Kohs's blocks, is presented as stimulus material for four progressive responses: (*a*) copying the top design, (*b*) copying the top design "with different colors," (*c*) copying the three-dimensional design, and (*d*) copying the three-dimensional design with a shift of colors. In view of the well-established findings that organic cases have difficulty with complex spatial perceptions, and also with ability to "shift," this double-barreled testing method rests on a sound rationale. Both an objective scoring system and qualitative indications are used to interpret the test. The retest reliability is reported as .85, and validity data on 276 cases show a clear separation of organics from normals and undeteriorated schizophrenics with little overlapping. Although the test needs further cross validation, it seems likely to become an interesting and efficient method for studying the psychological effects of brain pathology.—*L. F. S.*

Kuhlmann, F., & Anderson, Rose G. *Kuhlmann-Anderson Intelligence Tests.* (6th Ed.) 9 levels, ages 5-adult. 1 form at each level. 20-30 (40-45) min. Test booklet, each level ($2.40 per 25), with key, directions, pp. 42, and handbook, pp. 40; master manual, pp. 151 ($2.50); specimen set, 1 level ($1.00). Princeton, N. J. (188 Nassau St.): Personnel Press, 1952.

It is gratifying to see a useful psychological instrument kept up to date by continued research. Although the test materials are basically the same as those first published in 1927, new analyses of reliability, discriminative power, and norms have been made in 1951. Among group tests, the Kuhlmann-Anderson has always been distinguished by a clinical or "individual" flavor. It is not the easiest of such instruments to administer and score, but qualified users have felt rewarded for their efforts. Either timed or untimed administration is possible, with reliabilities centering around .95. The manual provides recently verified norms, and extensive statistical data including subtest intercorrelations. The new edition gives the impression that the test is by no means outdated, but is worthy of more years of service.—*L. F. S.*

Books Received

Allen, Robert M. *Introduction to the Rorschach technique.* New York: International Universities Press, 1953. Pp. 126. $3.00.

Crawford, Claude C., Cooley, Ethel G., Trillingham, C. C., & Stoops, Emery. *Living your life.* Boston: Heath, 1953. Pp. xv + 448. $2.80.

Devereux, George. (Ed.) *Psychoanalysis and the occult.* New York: International Universities Press, 1953. Pp. iv + 220. $3.00.

Field, Minna. *Patients are people.* New York: Columbia Univer. Press, 1953. Pp. viii + 244. $3.75.

Natenberg, Maurice. *Freudian psycho-antics.* Chicago: Regent House, 1953. Pp. 101. $2.00.

Piaget, Jean. *The origins of intelligence in children.* New York: International Universities Press, 1953. Pp. xi + 419. $6.00.

Podolsky, Edward. (Ed.) *Encyclopedia of aberrations.* New York: Philosophical Library, 1953. Pp. viii + 550. $10.00.

Reid, Joseph H., and Hagan, Helen R. *Residential treatment of emotionally disturbed children.* New York: Child Welfare League of America, Inc., 1952. Pp. ix + 313. $3.50.

Slavson, S. R. *Child psychotherapy.* New York: Columbia Univer. Press, 1952. Pp. xiii + 332. $4.50.

Tuckman, Jacob, and Lorge, Irving. *Retirement and the industrial worker.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1953. Pp. xvi + 105. $2.75.

The American Association of Social Workers

Publishes material of interest to social scientists as well as to professional social workers. Its quarterly *Social Work Journal* presents timely articles dealing with social welfare problems, administration of services, research, community relations, and similar topics. ($3.00 per year.) Its *Social Work Year Book*, a combined encyclopedia and directory of agencies, is the major reference work in its field. ($5.00 per copy.) Its *Membership Directory 1952* ($5.00) and its *International Welfare Directory 1952* ($1.00) are widely used. In 1952 the Association also published an Abridgment of *Social Work Education in the United States* by Hollis and Taylor (50 cents), *Social Workers in 1950* ($1.00), a reprint of *Common Human Needs* by Towle ($1.00), *Standards for the Professional Practice of Social Work* (25 cents), and numerous other monographs. Heading its 1953 list are *Education for Social Work Administration* by Trecker, Glick, and Kidneigh (50 cents) and *Report of a Study of School Social Work Practice in Twelve Communities* by Sikkema ($1.00).

◆ ◆ ◆

Write for a complete list of publications to

AMERICAN ASSOCIATION OF SOCIAL WORKERS

One Park Avenue **New York 16, N. Y.**

IN OCTOBER

WATCH FOR A NEW BOOK BY . . .

John Dollard, Frank Auld, Jr., Alice Marsden White

STEPS IN PSYCHOTHERAPY

This book on brief psychotherapy demonstrates psychiatric technique, using as illustrative material a detailed account of the handling of one therapeutic case by a student psychiatrist. Each page of the account is divided giving the detailed narrative of the student therapist's consultations with the patient at the top and the corresponding, critical comments of the supervisor at the bottom. Thus, the case is not simply reported; rather, the authors explain what happened and why in terms of psychological theory. Following the case there is a chapter discussing the sex-fear conflict in marriage.

One complete, hour-long interview transcribed verbatim from tape recording, and a transcript of the psychological tests given to the patient are notable features of the book. A glossary of the most important psychological terms used in the text is provided to aid readers not familiar with the technical terminology of the field.

The Macmillan Company

60 FIFTH AVENUE, NEW YORK 11, N.Y.